



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2017

---

## **Drosophila Nnf1 paralogs are partially redundant for somatic and germ line kinetochore function**

Blattner, Ariane C ; Aguilar-Rodríguez, José ; Kränzlin, Marcella ; Wagner, Andreas ; Lehner, Christian F

**Abstract:** Kinetochores allow attachment of chromosomes to spindle microtubules. Moreover, they host proteins that permit correction of erroneous attachments and prevent premature anaphase onset before bi-orientation of all chromosomes in metaphase has been achieved. Kinetochores are assembled from subcomplexes. Kinetochore proteins as well as the underlying centromere proteins and the centromeric DNA sequences evolve rapidly despite their fundamental importance for faithful chromosome segregation during mitotic and meiotic divisions. During evolution of *Drosophila melanogaster*, several centromere proteins were lost and a recent gene duplication has resulted in two Nnf1 paralogs, Nnf1a and Nnf1b, which code for alternative forms of a Mis12 kinetochore complex component. The rapid evolutionary divergence of centromere/kinetochore constituents in animals and plants has been proposed to be driven by an intragenome conflict resulting from centromere drive during female meiosis. Thus, a female meiosis-specific paralog might be expected to evolve rapidly under positive selection. While our characterization of the *D. melanogaster* Nnf1 paralogs hints at some partial functional specialization of Nnf1b for meiosis, we have failed to detect evidence for positive selection in our analysis of Nnf1 sequence evolution in the *Drosophila* lineage. Neither paralog is essential, even though we find some clear differences in subcellular localization and expression during development. Loss of both paralogs results in developmental lethality. We therefore conclude that the two paralogs are still in early stages of differentiation.

DOI: <https://doi.org/10.1007/s00412-016-0579-4>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-125398>

Journal Article

Accepted Version

Originally published at:

Blattner, Ariane C; Aguilar-Rodríguez, José; Kränzlin, Marcella; Wagner, Andreas; Lehner, Christian F (2017). *Drosophila Nnf1 paralogs are partially redundant for somatic and germ line kinetochore function*. *Chromosoma*, 126(1):145-163.

DOI: <https://doi.org/10.1007/s00412-016-0579-4>

## ***Drosophila Nnf1* paralogs are partially redundant for somatic and germline kinetochore function**

Ariane C. Blattner<sup>1)</sup>, José Aguilar-Rodríguez<sup>2,3)</sup>, Marcella Kränzlin<sup>1)</sup>, Andreas Wagner<sup>2,3,4)</sup>,  
Christian F. Lehner<sup>1,5)</sup>

1) Institute of Molecular Life Sciences (IMLS), University of Zurich, 8057 Zurich,  
Switzerland

2) Institute of Evolutionary Biology and Environmental Studies (IEU), University of Zurich,  
8057 Zurich, Switzerland.

3) Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland.

4) Santa Fe Institute (SFI), Santa Fe, New Mexico 87501, United States of America.

5) author for communication:

Institute of Molecular Life Sciences (IMLS)

University of Zurich

Winterthurerstrasse 190

8057 Zurich

Switzerland

e mail: christian.lehner@imls.uzh.ch

## Abstract

Kinetochore allow attachment of chromosomes to spindle microtubules. Moreover, they host proteins that permit correction of erroneous attachments and prevent premature anaphase onset before bi-orientation of all chromosomes in metaphase has been achieved. Kinetochore are assembled from subcomplexes. Kinetochore proteins as well as the underlying centromere proteins and the centromeric DNA sequences evolve rapidly despite their fundamental importance for faithful chromosome segregation during mitotic and meiotic divisions. During evolution of *Drosophila melanogaster*, several centromere proteins were lost and a recent gene duplication has resulted in two *Nnf1* paralogs, *Nnf1a* and *Nnf1b*, which code for alternative forms of a Mis12 kinetochore complex component. The rapid evolutionary divergence of centromere/kinetochore constituents in animals and plants has been proposed to be driven by an intragenome conflict resulting from centromere drive during female meiosis. Thus, a female meiosis-specific paralog might be expected to evolve rapidly under positive selection. While our characterization of the *D. melanogaster* *Nnf1* paralogs hints at some partial functional specialization of *Nnf1b* for meiosis, we have failed to detect evidence for positive selection in our analysis of *Nnf1* sequence evolution in the Drosophilid lineage. Neither paralog is essential, even though we find some clear differences in subcellular localization and expression during development. Loss of both paralogs results in developmental lethality. We therefore conclude that the two paralogs are still in early stages of differentiation.

**Keywords:** meiosis, centromere drive, Mis12 complex, gene duplication, kinetochore

## Introduction

Centromeres are essential for faithful chromosome segregation during mitotic and meiotic divisions. Nevertheless, centromeres evolve rapidly. Rapid divergence occurs in centromeric DNA sequences and also in the associated centromere and kinetochore proteins (Fukagawa and Earnshaw 2014; Henikoff et al. 2001). Centromeric DNA sequences can be short as in *Sacharomyces cerevisiae* where they conform to a 125 bp consensus sequence that is sufficient to specify centromere identity. In contrast, animal and plant species have centromeric DNA extending over many kilobases, either dispersed along holocentric chromosomes as in *Caenorhabditis elegans*, or focused into regional centromeres as for example in *Drosophila melanogaster* and *Homo sapiens*. Moreover, centromeric DNA in animals and plants is usually highly repetitive. These rapidly evolving tandem repeat sequences are of little immediate importance for centromere function, as clearly demonstrated by the identification and characterization of human neocentromeres (Marshall et al. 2008; Voullaire et al. 1993). Some of the neocentromeres that were identified in human patients appear to be fully functional even though they are located in regions devoid of any alpha-satellite repeats which are characteristically present within the normal centromeres of all human chromosomes. Epigenetic rather than DNA sequence-based centromere specification has been demonstrated experimentally in several organisms (Fukagawa and Earnshaw 2014; Karpen and Allshire 1997).

The fact that centromere-specific chromatin rather than a specific DNA sequence motif marks centromere identity in most eukaryotes explains the divergence of centromeric DNA sequences to some extent. However, it does not clarify immediately why centromere proteins are also strongly diverged (Meraldi et al. 2006; Schleiffer et al. 2012). For example, a most important centromere protein is the centromere-specific histone H3 variant named

Cenp-A in humans. While orthologs can be identified throughout the eukaryotic domain of life, their sequences are far less conserved than those of the canonical histone H3 proteins (Henikoff et al. 2000; Talbert et al. 2004). Moreover, several species including some kinetoplastids (Akiyoshi and Gull 2014; Berriman et al. 2005) and holocentric insects (Drinnenberg et al. 2014) have lost the centromere-specific histone H3 gene. This loss is even more surprising given that Cenp-A and the orthologous proteins of other species are assembled into centromere-specific nucleosomes that appear to function as epigenetic centromere mark (Fachinetti et al. 2013; Guse et al. 2011; Hori et al. 2013; Mendiburo et al. 2011; Westhorpe and Straight 2015). In addition, Cenp-A starts a recruitment cascade that allows assembly of kinetochore proteins at the start of M phase. As in the case of Cenp-A, the additional centromere and kinetochore proteins in general are characterized by limited sequence conservation and some lineage-specific gene losses. *D. melanogaster* provides a well-studied example illustrating the evolutionary plasticity of centromere and kinetochore organization. All 16 centromere proteins of the so-called CCAN (constitutive centromere associated network) with the exception of Cenp-C are absent in the fly (Heeger et al. 2005; Przewloka et al. 2007; Schittenhelm et al. 2007; Westermann and Schleiffer 2013). Cenp-C, which binds directly to Cenp-A/Cid, recruits the kinetochore proteins of the KMN network composed of Knl-1/Spc105 and the heterotetrameric Mis12- and Ndc80-complexes. Several features of the KMN network are distinct in *Drosophila*. *Drosophila* Spc105 has divergent MELT repeats even though canonical repeats have been shown to be essential for the binding of the spindle assembly checkpoint proteins in yeast and mammals (London et al. 2012; Primorac et al. 2013; Schittenhelm et al. 2009; Shepperd et al. 2012). Kmn2, the putative *Drosophila* Spc24 homolog (Schittenhelm et al. 2007), lacks the characteristic coiled-coil region present in other organisms where it is crucial for integration into the Ndc80 complex. Moreover, the *Drosophila* Mis12 complex lacks the Dsn1 subunit present in yeast and

vertebrates (Przewloka and Glover 2009; Przewloka et al. 2007; Schittenhelm et al. 2007). Overall, in comparison to mammals, *Drosophila* has evolved a far simpler centromere and kinetochore structure. The main microtubule binding activity at the kinetochore, the Ndc80/Nuf2 heterodimer, is linked to centromeric Cenp-A/Cid nucleosomes uniquely via Cenp-C and a variant KMN network, while other eukaryotes use CCAN components as an additional platform for KMN recruitment (Gascoigne et al. 2011; Kim and Yu 2015; Klare et al. 2015; Malvezzi et al. 2013; Nishino et al. 2013; Rago et al. 2015; Schleiffer et al. 2012).

The rapid divergence of both centromeric DNA sequences and associated proteins has been proposed to involve an evolutionary conflict in animals and plants (Henikoff et al. 2001; Malik and Henikoff 2009). The proposed intragenome conflict appears to arise in these species because the female sex transmits only one of the four meiotic haploid products to the next generation. Centromere sequence variants that induce their preferential meiotic segregation into the female pronucleus are predicted to increase their abundance in the population very rapidly. Selfish centromere variants might result from expansion of existing or new centromeric satellite arrays enabling enhanced recruitment of centromere proteins or spindle microtubules. Although centromeric DNA repeats are not absolutely essential for centromere function, it seems very likely that they exert some influence given their pervasive presence in established eukaryotic centromeres (Fukagawa and Earnshaw 2014). Spindle assembly during female meiosis is often acentrosomal, as also in *Drosophila*, where microtubules accumulate around meiotic chromosomes, thereby providing opportunities for selfish centromeres to bias their orientation within the asymmetric spindles via associated proteins that affect microtubule behavior. As the expansion of a selfish centromere within a population is likely to have negative effects, including retention of linked deleterious mutations, positive selection is predicted to favor centromere/kinetochore protein variants that specifically suppress the strength of an expanding selfish centromere (Henikoff et al.

2001; Malik and Henikoff 2009). Intriguingly, the centromere-specific histone H3 variants of *Drosophila*, primates, fish and plants, but not of budding yeast where meiosis is always symmetric, appear to have evolved under positive selection. Additional centromere proteins have been found to evolve under positive selection (Abbey and Kral 2015; Axelsson et al. 2010; Beck and Llopart 2015; Cooper and Henikoff 2004; Finseth et al. 2015; Malik and Henikoff 2001; Schueler et al. 2010; Talbert et al. 2004; Talbert et al. 2002; Yuan et al. 2015). Moreover, although still few, cases of centromere drive have been clearly documented in the plant *Mimulus* (Finseth et al. 2015; Fishman and Saunders 2008) and in the mouse (Chmatal et al. 2014).

Centromere drive might also be relevant for the retention of duplicated *Nnf1* gene copies within the *Drosophila melanogaster* subgroup (Schittenhelm et al. 2007). *Nnf1* (Necessary for nuclear function 1) was originally identified in budding yeast and subsequently shown to be a component of the Mis12 complex (Shan et al. 1997; Westermann et al. 2003). The Mis12 complex subunits Nsl1 and Dsn1 form direct contacts with Knl-1/Spc105 and the Spc24/25 subunits of the Ndc80 complex (Malvezzi et al. 2013; Petrovic et al. 2014; Petrovic et al. 2010). By binding to Cenp-C, the Mis12 complex recruits the KMN network to the centromere (Przewloka et al. 2011; Screpanti et al. 2011). *Drosophila* *Nnf1* binds directly to Cenp-C in vitro (Przewloka et al. 2011). Crosslinking analyses in yeast have revealed additional interactions of Mis12 complex subunits and Cenp-C (Hornung et al. 2014).

An initial comparison of the expression pattern of the two paralogous *Nnf1* genes in *D. melanogaster* (Schittenhelm et al. 2007) suggested that *Nnf1b* might be germline-specific, raising the possibility that it provides a meiosis-specific function. Accordingly, this but not the other paralog might evolve under positive selection when engaged in suppression of a hypothetical centromere drive during female meiosis. To address these possibilities we have

further characterized the two *Drosophila NnfI* paralogs. Beyond evolutionary sequence analyses, we have generated null mutations and report single and double mutant phenotypes. Our analyses reveal a partial functional specialization of the *NnfI* paralogs without evidence for strong positive selection.



## Materials and methods

### Drosophila genetics

*Df(2R)Exel6070* and *Df(2R)Exel7164* (Parks et al. 2004) were obtained from the Bloomington Drosophila Stock Center. PCR assays using the primer pair AB3 (5'-CGTCCAAGCTGTTCGCTAGT-3') and AB4 (5'-AACGGCTTTTTGCGTAATTG-3') were used for confirmation that *Df(2R)Exel7164* deletes *Nnf1a*. These primers anneal to regions flanking the deficiency breakpoint and amplify a 2 kb fragment only when the deficiency is present in genomic template DNA.

*PBac{SAstopDsRed}<sup>LL02791</sup>* (Schuldiner et al. 2008) was obtained from the Drosophila Genetic Resource Center (DGRC, Kyoto Institute of Technology). Information on the transposon insertion site provided by FlyBase and DGRC were not fully consistent. Our analyses by PCR and sequencing revealed an insertion site in the second TTAA motif within intron 1 of *Nnf1a*.

Two deficiencies that delete *Nnf1b* were isolated. *Df(2L)MK01* was generated as described (Parks et al. 2004) by Flp-mediated recombination between the two FRT transposon insertions *P{XP}d00346* and *PBac{WH}f01300* (Thibault et al. 2004) which were obtained from the Exelixis Collection at the Harvard Medical School. For the isolation of *Df(2L)NK01*, we mobilized the transposon insertion *P{SUPor-P}KG07276* (Bellen et al. 2004) (Bloomington Drosophila Stock Center, #14327), which is marked with a mini-*w*<sup>+</sup> gene, by crossing with *y<sup>l</sup> w<sup>\*</sup>; CyO, H{w<sup>+</sup>mC=PDelta2-3}HoP2.1/Bc* (Bloomington Drosophila Stock Center, #2078). Single white-eyed progeny males obtained from a cross of *w<sup>\*</sup>; P{SUPor-P}KG07276/CyO, H{w<sup>+</sup>mC=PDelta2-3}HoP2.1* males with *w<sup>\*</sup>; If/CyO* females were used to establish 250 lines after backcrossing to *w<sup>\*</sup>; If/CyO* females. None of the *w<sup>-</sup>* revertant

second chromosomes present in these lines was free of recessive lethal mutations, suggesting the presence of such second site mutations on the starting *P{SUP<sup>or</sup>-P}KG07276* chromosome. Genomic DNA isolated from balanced *w<sup>\*</sup>* revertant flies was analyzed with a multiplex PCR using the three primers OZH47 (5'-TCGCACCACAAAAAGTCAAC-3'), OZH48 (5'-CCTCCAGGAACAGAAGCAGA-3'), and CL18 (5'-CGCAGGTACCACCTTATGTTATTTTCATCATG-3') for the identification of imprecise excision events. These primers amplify three fragments of different length from *P{SUP<sup>or</sup>-P}KG07276/CyO*, one from the balancer and two from the transposon chromosome which include the regions flanking the transposon insertion on the left and right side, respectively. Revertant lines, where specifically the flanking fragment on the *Nnf1b* side could no longer be amplified, were further characterized. In case of *Df(2L)NK01*, additional characterization by PCR and sequencing revealed the presence of a deletion with breakpoints within the *P{SUP<sup>or</sup>-P}* transposon just upstream of *Nnf1b* and within the second exon of the *Dbp2IE2* gene just downstream of *Nnf1b*. To confirm the absence of *Nnf1b* gene in *w<sup>\*</sup>*; *Df(2L)MK01*, *gDSP II.2/Df(2L)NK01*, *gDSP II.2* flies, a multiplex PCR with the *Nnf1b*-specific primers NT60 (5'-CTGAGGTTCTGATGGGTGGT-3') and NT53 (5'-TTAATGCTTTTGGCCATAGCAA-3') in combination with the primers AB1 (5'-TTCAATGGAGGATTCGGAAG-3') and NT58 (5'-GCTTGTCATGGACATGTTGG-3') which amplify a control fragment from *Nnf1a* were used. In combination with wild-type genomic DNA this multiplex PCR results in the amplification of a 826 bp *Nnf1a* and a 388 bp *Nnf1b* fragment.

For *Nnf1a Nnf1b* double mutant analysis we generated stocks with second chromosomes obtained by standard meiotic recombination: *w<sup>\*</sup>*; *Df(2L)MK01*, *gDSP II.2*, *PBac{SAstopDsRed}<sup>LL02791</sup>/CyO* and *w<sup>\*</sup>*; *Df(2L)NK01*, *gDSP II.2*, *Df(2R)Exel7164/CyO*. 25% of the zygotes generated by a cross between these two stocks have the genotype *w<sup>\*</sup>*;

*Df(2L)MK01*, *gDSP II.2*, *PBac{SAstopDsRed}<sup>LL02791</sup>/Df(2L)NK01*, *gDSPII.2*, *Df(2R)Exel7164* and therefore neither zygotic *Nnf1a* nor *Nnf1b* function. Stocks with a *CyO* version carrying *P{Dfd-GMR-nvYFP}* (Le et al. 2006) were used for the identification of double mutant embryos and larvae based on absence of EYFP fluorescence in the head region.

Lines carrying the *gDSP*, *g-Nnf1a* and *g-Nnf1b* transgenes were generated by standard germline transformation using the pCaSpeR-4 constructs described below. Lines carrying the transgenes *g-EGFP-Nnf1a*, *g-Nnf1a-EGFP*, *g-EGFP-Nnf1b*, *g-Nnf1b-EGFP* were obtained by integrating the *attB* constructs described below into the *attP* landing site *PBac{y<sup>+</sup>-attP-9A}/VK00020* (Venken et al. 2006). Lines carrying the transgenes *ga-EGFP-Nnf1a* and *ga-EGFP-Nnf1b* were made by integrating the *pattB* constructs described below into the *attP* landing site *P{CaryP}attP2* (Groth et al. 2004).

Crosses for phenotypic analyses were performed at 25°C. *w<sup>l</sup>* was used as wild-type control. To assess male fertility, single males of a given genotype were crossed to three *w<sup>l</sup>* virgin females. For examination of female fertility, three virgins of a genotype were pooled and crossed to three *w<sup>l</sup>* males. In both cases, ten replicate crosses were set up. Flies were allowed to mate for two days, then transferred to a fresh vial and discarded after two more days. The eclosing adult progeny from the second vial was counted for eight days. For X chromosome nondisjunction tests (X-ND) we used the stock *C(1;Y)I, y v f B: y<sup>+</sup>/C(1)RM, y<sup>2</sup> su(w<sup>a</sup>) w<sup>a</sup>* (Bloomington Drosophila Stock Center, # 700). Males with a compound XY chromosome were crossed to *Nnf1b<sup>null</sup>* females. Regular gametes produced by these females develop into either XXY females expressing *Bar* eyes or X0 males with normal eyes after fertilization with XY or 0 sperm, respectively. Irregular nullo-X gametes produced by females after X-ND develop into XY males with *Bar* eyes after fertilization with XY sperm. Moreover, irregular diplo-X gametes resulting from X-ND in females develop into adult

females with normal eyes after fertilization with 0 sperm. In contrast, XXXY zygotes resulting after fertilization of irregular diplo-X oocytes with XY sperm, as well as 00 zygotes resulting after fertilization of irregular nullo-X oocytes with 0 sperm do not develop to the adult stage. Therefore, to determine the rate of X-ND, we multiplied the number of irregular adult progeny (males with *Bar* eyes and females with normal eyes) by two before dividing by the total number of adult progeny.

### Plasmids

A pCaSpeR-4 construct was made for the generation of *gDSP* transgenic flies. The pCaSpeR4-gDSP construct contains a 6.3 kb genomic fragment with the genes *Dbp21E2*, *Saf6* and *Pex12* that are deleted in *Df(2L)MK01* apart from *Nnf1b*. The first part of this genomic region was amplified using the primers OZH-20 (5'-ATATGGTACCCTTCGATTTTAGATTGCTATGGCC-3') and OZH-21 (5'-ATATCCCGGGGTTAGCCTGGATCCTTATCTTTGG-3') from BAC16I01 (Hoskins et al. 2000) which contains an insert fragment of the corresponding *Drosophila* genome region. After digestion with *Acc65I* and *SmaI*, the fragment was inserted into the corresponding sites of pSLfa1180fa (Horn et al. 2000). The resulting first cloning intermediate was then digested with *PacI* and *SmaI* for insertion of the second part of the genomic region which was isolated from the same BAC with primers OZH-22 (5'-CCGCGCACTGCTCTACTGATATACC-3') and OZH-23 (5'-ATATCCCGGGCCGGTCCGAGGTTTCGGCTGCC-3') and digested with the same restriction enzymes. The complete 6.3 kb genomic region was released with *Acc65I* and *SmaI* from the second cloning intermediate and inserted into the *Acc65I* and *HpaI* sites of pCaSpeR-4.

We also generated pCaSpeR-4 constructs for the production of *g-Nnf1b* and *g-Nnf1a* transgenic flies. Primers AB26 (5'-TAAAATCGCATCTTGCTTG-3') and AB31 (5'-CAATTCTAGACATTTGCCGGCAAGT-3') were used for the enzymatic amplification of the *Nnf1b* genomic region from pFlyFos030346 (Ejsmont et al. 2009). After digestion with BamHI and XbaI, the fragment was inserted into the corresponding sites of pCaSpeR-4. A 1.8 kb fragment with *Nnf1a* was amplified from *w<sup>1</sup>* genomic DNA using the primers CL126 (5'-TTTGGCGGCCGCTAAACACTGGCTGGTAAAATAT-3') and CL127 (5'-GCAATGGTACCAATTTAAATAAATTGTGAATGTC-3'). After digestion with NotI and Acc65I, the fragment was inserted into the corresponding sites of pCaSpeR-4. Characterization of construct and *w<sup>1</sup>* genomic DNA by sequencing revealed some genetic polymorphisms where the *w<sup>1</sup>* genome sequence differs from the FlyBase genome sequence.

For the generation of attB transgene constructs allowing expression of *Nnf1a* and *Nnf1b* with N- or C-terminal EGFP extension, we first derived the vectors pattB-ORF-EGFP and pattB-EGFP-ORF from pattB (Bischof et al. 2013). In case of pattB-ORF-EGFP, the primers AB17 (5'- GACTGCGGCCGCAATGGTGAGCAAGGGC-3') and AB19 (5'-GACTCTCGAGTCACTTGTACAGCTCGTCCATG-3') were used for enzymatic amplification of the EGFP coding region (with stop codon), followed by digestion with NotI and XbaI and insertion into the corresponding sites in pattB. In case of pattB-EGFP-ORF, the primers AB17 and AB18 (5'-GACTCTCGAGCTTGTACAGCTCGTCCATG-3') were used for enzymatic amplification of the EGFP coding region (without stop codon) followed by digestion with NotI and XbaI and insertion into the corresponding sites in pattB.

For the construction of pattB-g-Nnf1a-EGFP, the 3' flanking region of *Nnf1a* was first amplified with primers AB24 (5'-GAAGCTCGAGACGACTTCTGACAACTTAAATG-3') and AB25 (5'-GACTTCTAGAAATTTAAATAAATTGTGAATGTCAATG-3') from pCaSpeR-4-g-Nnf1a and cloned into pattB-ORF-EGFP using XhoI and XbaI. The *Nnf1a*

gene and upstream sequences were then inserted using BamHI and NotI after enzymatic amplification with the primers AB20 (5'-GACTGGATCCCTAAACACTGGCTGGTAAAATATT-3') and AB23 (5'-TTAAGCGGCCGCGAAGTCGTTCAATGCTTCG-3').

For the construction of pattB-g-EGFP-Nnf1a, the *Nnf1a* gene and 3' flanking sequences were amplified from pCaSpeR-4-g-Nnf1a with the primers AB22 (5'-AATTCTCGAGATGGAGGATTCGGAAGC-3') and AB25 (5'-GACTTCTAGAAATTTAAATAAATTGTGAATGTCAATG-3') and cloned into the corresponding sites of pattB-EGFP-ORF using XhoI and XbaI. The *Nnf1a* upstream sequences were then inserted using BamHI and NotI after enzymatic amplification with the primers AB20 and AB21 (5'-CGAAGCGGCCGCTGAAATAATTTGTTAGTAAATTCAAATT-3').

For the construction of pattB-g-Nnf1b-EGFP, the *Nnf1b* 3' flanking region was amplified from *w<sup>1</sup>* genomic DNA with the primers AB30 (5'-TAAACTCGAGGTATGTGCAGGAAGAAATGTAA-3') and AB31 (5'-CAATTCTAGACATTTGCCGGCAAGT-3') and cloned into pattB-ORF-EGFP using XhoI and XbaI. The *Nnf1b* gene and 5' upstream sequences were then amplified with the primers AB26 (5'-TAAAATCGCATCTTGCTTG-3') and AB29 (5'-TTTAGCGGCCGCCATTTCTTCCTGCACATACG-3') and inserted using BamHI and NotI.

For the construction of pattB-g-EGFP-Nnf1b, the *Nnf1b* gene and 3' flanking sequences were amplified from pFlyFos030346 with AB28 (5'-TGTTGTCGACATGAATAATATTGAAGAGGACACC-3') and AB31 (5'-CAATTCTAGACATTTGCCGGCAAGT-3'), digested with SalI and XbaI followed by cloning into XhoI/XbaI-digested pattB-EGFP-ORF. The *Nnf1b* upstream sequences were then amplified with AB26 and AB27 (5'-

AATAGCGGCCGCACTTTTAACACAATTCTGGCA-3') and inserted using BamHI and NotI.

For the construction of pattB-ga-EGFP-Nnf1b, we first removed the region coding for EGFP-Nnf1a and the 3' flanking region from pattB-g-EGFP-Nnf1a using NotI and XbaI. The deleted region was then replaced with a fragment where EGFP-Nnf1b is fused to Nnf1a 3' sequences obtained by enzymatic amplification from pattB-g-EGFP-Nnf1b using the primers AB17 and AB90 which is 199 bp long and introduces the Nnf1a 3' sequences (UTR and flanking region).

The correctness of the transcribed region in all the final attB constructs was confirmed by sequencing.

## qRT-PCR

Total RNA was isolated from 0-2 hour embryos collected from either

*PBac{SAstopDsRed}<sup>LL02791</sup>*, *PBac{SAstopDsRed}<sup>LL02791</sup>/CyO, w\**; *Df(2L)MK01/Df(2L)NK01*; *gDSP III.2/+*, or *w<sup>l</sup>* control flies. 300-1500 embryos for each genotype were used for RNA isolation with TRIzol (Invitrogen) followed by DNase digestion (DNA-free Kit, Ambion). cDNA synthesis was performed with 500 ng RNA per reaction using a Transcriptor High Fidelity cDNA Synthesis Kit (Roche). Real-time quantitative PCR was performed using SYBR Green with an Applied Biosystems 7900HT using the recommended two-step cycling protocol with the following primer pairs: AB1 (5'-TTCAATGGAGGATTCGGAAG-3') and NT46 (5'-TCAAGTCGGCAAAGATTTGTT-3'), NT47 (5'-AACAAATCTTTGCCGACTTGA-3') and NT48 (5'-TGAGGGTTAATTGCGTCACT-3'), NT47 and AB2 (5'-TCTACGATGCTTTCGGTGTC-3'), NT50 (5'-

CATGCGATTTTGGACTTCA-3') and NT51 (5'-GCGATTGGCATTGCTCTT-3'), NT57 (5'-GACGTGGAGACCAAGCTGAC-3') and NT61 (5'-TCCTCAGGGGACAGACAGTT-3'). For normalization, three primer pairs (sequences provided upon request) were used for amplification of transcripts from the genes *Act5C*, *alphaTub84B* and *Tbp*. At least three replicates were performed.

### **Immunofluorescence and immunoblotting**

Eggs were collected from crosses that generate *Nnf1a Nnf1b* double mutants on apple agar plates at 25°C for 3 hours followed by ageing for 13 hours before fixation with 4% formaldehyde in phosphate buffered saline (PBS) and devitellinization. For analyses of embryos during the syncytial blastoderm stages, we collected eggs for 1 hour followed by ageing for an additional hour before fixation and devitellinization. Embryos were stained with a rabbit antiserum against *Drosophila* Cyclin B (Jacobs et al. 1998) (1: 2000) and a DNA stain (Hoechst 33258, 1 µg/ml) according to standard procedures. Ovaries and testes were dissected in testis buffer (183 mM KCl, 47 mM NaCl, 10 mM Tris-HCl pH 6.8) and fixed in 4% formaldehyde in PBS, 0.1% Triton X-100 for 10 min. DNA was stained with Hoechst 33258 (1 µg/ml) in PBS, 0.1% Triton X-100 for 10 min. Samples were washed twice with PBS and mounted in 70% Glycerol, 50 mM Tris-Cl pH 8.5, 10 mg/ml propyl gallate, 0.5 mg/ml phenylenediamine. Testis squash preparations were prepared essentially as described previously (White-Cooper 2004). For quantification of the kinetochore signals in syncytial embryos, image stacks of 15 focal planes with 300 nm spacing that contained the nuclear layer were acquired with a 100x/1.4 objective and a Cell Observer HS (Zeiss) microscope. For quantification of the kinetochore signals in prometaphase I cells in testes, focal planes



with 250 nm spacing were acquired with a 63x/1.4 objective. After maximum projection, kinetochore signals were quantified as described previously (Schittenhelm et al. 2010). Maximum intensity projections are displayed in the figures except where stated differently.

Immunoblotting was performed using anti-EGFP (Schittenhelm et al. 2007) at 1:4000 and anti-PSTAIR (Sigma, P7962) at 1:4000. Peroxidase-coupled goat antibodies (Jackson Immuno Research) against mouse or rabbit IgG were used as secondary antibodies at 1:2000 and detected using Amersham ECL (GE Healthcare). Quantification of band intensities was performed as previously described (Radermacher et al. 2014).

### **Phylogenetic reconstruction**

We obtained the coding DNA sequences of *Nnf1* from nineteen different *Drosophila* species using BLAST searches of their annotated genomes. Nine species predate the duplication of *Nnf1* (*D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi*, *D. kikkawai*, *D. pectinata*). In the other ten species, *Nnf1* has been duplicated and their genomes harbor two copies of the gene (*D. simulans*, *D. sechellia*, *D. melanogaster*, *D. yakuba*, *D. erecta*, *D. eugracilis*, *D. biarmipes*, *D. takahashii*, *D. elegans*, *D. rhopaloea*).

We aligned the twenty-nine *Nnf1* coding sequences using the PRANK algorithm (Löytynoja and Goldman 2008), and employed a recent whole genome phylogeny as a guide tree for the aligner (Seetharam and Stuart 2013). We removed all nucleotide sites with gaps from the resulting codon-based nucleotide alignment using trimAl v1.2 (Capella-Gutierrez et al. 2009), which resulted in an alignment with 507 aligned nucleotides (or 169 aligned codons) (Online Resource 1, Supplementary Fig. 1). We then used MEGA v6.06 to calculate the mean pairwise nucleotide and amino acid distances between the paralogs *Nnf1a* and *Nnf1b* (Tamura

et al. 2013). To compute nucleotide distances we used the maximum likelihood composite model (Tamura et al. 2004), while for amino acid distances we used a Poisson correction (Zuckerkandl and Pauling 1965). We then used PhyML v3 (Guindon et al. 2010) with 1000 bootstrap replicates to build a gene tree through phylogenetic inference by maximum likelihood. Following the Akaike information criterion in jModelTest v2.1.7, we chose the nucleotide substitution model TIM1 (“transition model”) + I +  $\Gamma$  (Posada 2008). We rooted the tree according to the known phylogeny for *Drosophila* species (Seetharam and Stuart 2013). All downstream analyses of selection are based on the gene tree resulting from this procedure.

### **Analyses of selection**

To explore the selective pressures in the different lineages of the *Nnf1* tree, we estimated the ratio  $\omega$  of the nonsynonymous and synonymous substitution rates ( $\omega = d_N/d_S$ ). To perform these analyses we employed three classes of codon substitution models, the branch models, the sites models, and the branch-site models that we implemented using the codeml program from the package PAML v4.7 (Yang 2007). For all three models we used likelihood ratio tests (LRTs) to compare a complex model against a simpler nested model (Huelsenbeck and Crandall 1997; Yang 1998). We computed twice the difference in the log-likelihoods ( $2\Delta l$ ) of the two nested models and compared its value against a  $\chi^2$  distribution, with the degrees of freedom being the difference in the number of estimated parameters between the two models. In all these analysis, we estimated the codon frequencies empirically, using the codon model F3x4. Also, we estimated all branch lengths of the gene tree from the initial values obtained

with the one-ratio M0 model (Goldman and Yang 1994). This simple model assumes a single ratio  $\omega$  for all tree branches and all codon sites in the sequence alignment.

In our phylogenetic analysis, we first used *branch* models (Yang 1998), which allow us to estimate different  $\omega$  ratios for different branches in the phylogeny. However, for a given branch, these models do not allow variation in  $\omega$  among different codon sites. Second, we used *sites* models (Nielsen and Yang 1998; Wong et al. 2004; Yang et al. 2000; Yang et al. 2005), which do allow  $\omega$  to vary among sites but not among branches. We used these models to test if specific sites were affected by positive selection during the evolution of *Nnfl*.

Among these kinds of models, the neutral model M1a specifies two classes of codon sites: conserved sites under purifying selection ( $\omega < 1$ ) and neutral sites ( $\omega = 1$ ). We compared this model against the M2a model, which allows an additional class of codon sites that are under positive selection ( $\omega > 1$ ) (Nielsen and Yang 1998; Wong et al. 2004; Yang et al. 2000). We also compared the neutral model M7, where  $\omega$  varies among codons according to a beta distribution, to the model M8, which includes an additional codon class subjected to positive selection (Yang et al. 2000). Finally, we also used *branch-site* models (Yang and Nielsen 2002; Zhang et al. 2005), which allowed us to ask whether particular sites evolve under positive selection for specific branches (*foreground* branches, while the others are named *background* branches). For the branch and branch-site models, we divided the gene tree of *Nnfl* in pre-duplication branches (assumed to evolve at a rate  $\omega_0$ ), branches immediately following the duplication event ( $\omega_1$ ), and branches of the clades *Nnfla* ( $\omega_2$ ) and *Nnflb* ( $\omega_3$ ).

## Results

### *Nnflb* is not essential

The two *Nnfl* paralogs of *Drosophila melanogaster* appear to have arisen after duplication of a primordial *Nnfl* gene during evolution of the *melanogaster* species group presumably about 10 million years ago (Schittenhelm et al., 2007; see also below). The genomic location of the *Nnfla* paralog on chromosome 2R corresponds to that of the primordial *Nnfl* gene according to gene synteny analysis, while *Nnflb* is in a distinct region on chromosome 2L. The exon-intron structure of the two *Nnfl* paralogs is identical, indicating that the duplication event did not involve an mRNA intermediate. Published data concerning the expression pattern of *Nnflb* during *Drosophila* development indicate maximal transcript levels in ovaries of adult females and at the earliest embryonic stages (Chintapalli et al. 2007; Graveley et al. 2011; Schittenhelm et al. 2007). In contrast, *Nnfla* transcript levels are minimal at the earliest embryonic stages, but they increase later during embryogenesis when *Nnflb* transcripts rapidly vanish. Therefore, *Nnflb* appears to be the predominant form present during female meiosis, and the fixation of the *Nnflb* paralog during evolution might reflect functional specialization for support of meiosis-specific kinetochore behavior. The success of meiosis relies on specific control of sister kinetochore autonomy. At the onset of the first meiotic division, the two sister kinetochores behave as a functional unit. Their co-orientation in the meiosis I spindle allows for reductional segregation of homologs to opposite poles. In contrast, during all other divisions (meiosis II and mitosis), sister kinetochores behave as autonomous units, allowing their bi-orientation within the spindle. The control of sister kinetochore behavior during meiosis is poorly understood and a functional specialization of *Nnflb* in this context would be of great interest. To address *Nnflb* function, we generated zygotes lacking *Nnflb* for phenotypic analyses.

For elimination of *Nnflb*, we first generated a small deficiency, *Df(2L)MK01*, using Flp/FRT-mediated recombination between flanking transposon insertions carrying FRT sites

(Parks et al. 2004). Beyond *Nnf1b*, *Df(2L)MK01* deletes the three downstream genes *Dbp21E2*, *Saf6* and *Pex12* (Fig. 1a). To restore these three genes, we generated a transgene (*gDSP*) carrying an appropriate genomic fragment (Fig. 1a). In addition, we isolated *Df(2L)NK01*, an even smaller deficiency, as imprecise excision after mobilization of a P element insertion upstream of *Nnf1b* (Fig. 1a). This second deficiency deletes *Nnf1b* completely and *Dbp21E2* partially (Fig. 1a). To assess the loss-of-*Nnf1b* phenotype, we arranged a cross generating the genotype *Df(2L)MK01, gDSP II.2/Df(2L)NK01* which does not have any *Nnf1b* gene copy and only one functional *Dbp21E2* copy, but otherwise has two functional copies of all other genes in the region. For simplicity, we will designate this genotype as *Nnf1b<sup>null-1</sup>*. In addition, we also generated *Df(2L)MK01, gDSP II.2/Df(2L)NK01, gDSP II.2*, designated as *Nnf1b<sup>null-2</sup>*, which does not have *Nnf1b<sup>+</sup>* gene copies but an extra copy of both *Saf6<sup>+</sup>* and *Pex12<sup>+</sup>* compared to normal diploid wild-type.

*Nnf1b<sup>null</sup>* flies developed to the adult stage with an efficiency comparable to that of the *Nnf1b<sup>+</sup>* siblings (Fig. 1b,c). PCR assays confirmed the absence of the *Nnf1b* gene (Fig. 1b) and *Nnf1b* transcripts (Fig. 2b) in *Nnf1b<sup>null</sup>* flies. To assess the fertility of *Nnf1b<sup>null</sup>* flies, males and females were crossed with control flies followed by progeny counting. Complete absence of a meiosis-specific kinetochore protein is expected to result in very high levels of aneuploid progeny and consequentially in almost complete infertility. Unexpectedly, fertility of *Nnf1b<sup>null</sup>* flies was essentially normal in both sexes (Fig. 1d). However, these results do not exclude that absence of *Nnf1b* might cause a partial kinetochore defect. Meiotic chromosome segregation errors might therefore be increased in *Nnf1b<sup>null</sup>*, but only to a level that does not yet cause a substantial fertility depression. To assess the fidelity of meiosis in *Nnf1b<sup>null</sup>* flies we analyzed X chromosome nondisjunction (X-ND) in females. X-ND in *Nnf1b<sup>null</sup>* females was comparable to that of our control females (Fig. 1e). We conclude that *Nnf1b* is not absolutely essential for development into fertile adults.

### ***Nnf1a* is not essential**

The apparent absence of abnormalities resulting from complete elimination of *Nnf1b* gene function might reflect redundancy. To address potential functional overlap with *Nnf1a*, we determined whether the transposon insertion *PBac{SAstopDsRed}<sup>LL02791</sup>* eliminates *Nnf1a* gene function (Fig. 2a). The transposon was found to be integrated within the first intron. As the transposon includes a splice acceptor site followed by stop codons in all three frames (Schuldiner et al. 2008) it is expected to cause premature termination after translation of only the N-terminal 17% of the normal full length protein. To evaluate whether the splice acceptor site within the transposon is used indeed as expected, we performed quantitative RT-PCR experiments (Fig. 2b). As expected, in transposon heterozygotes we detected wild-type transcripts as well as transcripts containing the transposon derived exon. In contrast, in transposon homozygotes we detected exclusively the latter transcripts, and in wild-type controls only the former transcripts. We conclude that the transposon insertion precludes expression of normal transcripts and therefore, it appears to be a null allele.

To assess phenotypic consequences resulting from elimination of *Nnf1a*, we crossed balanced females heterozygous for *PBac{SAstopDsRed}<sup>LL02791</sup>* with balanced males of the same genotype, as well as with balanced males heterozygous for *Df(2R)Exel6070* or *Df(2R)Exel7164* which both delete *Nnf1a*. Adult progeny homozygous or hemizygous for *PBac{SAstopDsRed}<sup>LL02791</sup>* were obtained with frequencies very close to their Mendelian ratio (Fig. 2c), indicating that their viability is similar to that of the balanced siblings. Compared to hemizygotes, the homozygous flies were weaker and had a shorter life span, arguing for some second site effects in *PBac{SAstopDsRed}<sup>LL02791</sup>* homozygotes (Online

Resource 1, Supplementary Fig. 2). However, also hemizygous flies had a life span that was only about half of that of control flies. To evaluate the importance of *Nnfla* for fertility, we crossed homozygous and hemizygous *PBac{SAstopDsRed}<sup>LL02791</sup>* flies with *w<sup>1</sup>* control flies. Based on progeny counts, hemizygous males lacking *Nnfla* were fully fertile. Female fertility appeared to be at most slightly reduced in *PBac{SAstopDsRed}<sup>LL02791</sup>/Df(2R)Exel7164* flies, as the observed reduction was not statistically significant. The somewhat stronger effects observed in case of *PBac{SAstopDsRed}<sup>LL02791</sup>/Df(2R)Exel6070* and even more in case of *PBac{SAstopDsRed}<sup>LL02791</sup>* homozygotes presumably reflect genetic background effects. Similarly, X-ND was at most slightly elevated in hemizygous females compared to controls and a bit more in *PBac{SAstopDsRed}<sup>LL02791</sup>* homozygous females (0.3% in *w* control, 0.3% in *PBac{SAstopDsRed}<sup>LL02791</sup>/Df(2R)Exel6070*, 0.5% in *PBac{SAstopDsRed}<sup>LL02791</sup>/Df(2R)Exel7164*, 2.2% in *PBac{SAstopDsRed}<sup>LL02791</sup>*, *n* > 600). We conclude that *Nnfla* is not absolutely essential for development into adults, but vitality and perhaps to a minor extent also female fertility are no longer fully normal.

### ***Nnfla* and *Nnflb* have overlapping functions**

The development of fertile adults in the absence of either *Nnfla* or *Nnflb* function might be explained by functional overlap among the two paralogs. To address redundancy, we analyzed double mutants (*Df(2L)MK01*, *gDSP II.2*, *PBac{SAstopDsRed}<sup>LL02791</sup>/Df(2L)NK01*, *gDSP II.2*, *Df(2R)Exel7164*). Double mutants did not develop to the adult stage. The large majority of the double mutants (between 80-97%, *n* > 50) still completed embryogenesis and reached the early larval stages presumably by exploiting maternally provided Nnfl. Analysis of double mutant embryos revealed that they were no longer fully normal at late embryonic

stages during which abnormalities were clearly detectable within the central nervous system (CNS) (Fig. 3). During wild-type embryogenesis, mitotic cell proliferation lasts longer in the CNS than in most other embryonic tissues. The maternal *Nnfl* contribution therefore seems to run out before completion of the late CNS proliferation. The abnormalities observed in the double mutant CNS after labeling with a DNA stain and anti-Cyclin B were highly reminiscent of those observed in mutant embryos lacking zygotic function of other KMN components (Schittenhelm et al. 2009; Schittenhelm et al. 2007). Some cells were larger and displayed increased Cyclin B labeling, presumably reflecting cell cycle defects triggered by partially functional kinetochores and hyperploidy resulting from chromosome missegregation. These developmental abnormalities might perhaps also explain why double mutant larvae failed to grow and reach the third instar wandering stage. In conclusion, the complete developmental lethality of double mutants in contrast to the viability and fertility of single mutants indicates that the two *Nnfl* paralog have partially overlapping functions.

To establish that the developmental lethality of double mutants results from a loss of *Nnfl* function and not from genetic background effects, we performed rescue experiments with transgenes expressing *Nnfla* or *Nnflb* with N- or C-terminal EGFP extensions. The regulatory regions controlling expression of these transgenes were those present at the endogenous loci. All transgenes were inserted into the same chromosomal attP landing site.

When double mutants inherited either *g-EGFP-Nnfla* or *g-Nnfla-EGFP* from the father, they developed into fertile adults. These observations demonstrate that the lethality of double mutants reflects a loss of *Nnfl* function. Moreover, EGFP-*Nnfla* and *Nnfla*-EGFP are functional proteins.

In contrast to the *Nnfla* transgenes, *g-EGFP-Nnflb* and *g-Nnflb-EGFP* did not prevent the lethality of double mutants. However, interpretation of these rescue failures was complicated by our subsequent finding that a single endogenous *Nnflb*<sup>+</sup> copy does not



support the development of *Nnf1a* mutants to the adult stage and that homozygosity of the transgenes *g-EGFP-Nnf1b* and *g-Nnf1b-EGFP* resulted in lethality even in an otherwise wild-type background. Moreover, an additional, P element-based transgene (*g-Nnf1b*) with the same regulatory region driving expression of untagged Nnf1b failed to promote double mutant development to the adult stage even when present in two copies. Both *g-Nnf1b* insertions used in these experiments were homozygous viable and fertile in the wild-type background. We assume therefore that the *Nnf1b* regulatory regions present in our transgenes are not fully functional. Some regulatory elements might be outside of the 1.2 kb upstream region used in our transgenes, perhaps within the additional 20 kb of intergenic region upstream of *Nnf1b*.

### **Stage-specific differences in Nnf1a and Nnf1b subcellular localization during the cell cycle**

For a comparison of potential functional specialization of the paralogous Nnf1 proteins, we generated two additional transgenic lines (*ga-EGFP-Nnf1a* and *ga-EGFP-Nnf1b*). Both transgenes were inserted into the same chromosomal attP landing site and contained the same *Nnf1a* 5' and 3' regulatory region. However, they coded for either EGFP-Nnf1a or EGFP-Nnf1b. When present in one copy both transgenes supported the development of double mutants into fertile adults that eclosed with the expected Mendelian frequency (67% expected; 65% and 68% observed for *ga-EGFP-Nnf1a* and *ga-EGFP-Nnf1b*, respectively;  $n > 200$ ). Moreover, rescued females had at most slightly elevated X-ND ( $w$  control 0.3%; 0.4% and 0.5% for double mutants rescued with *ga-EGFP-Nnf1a* and *ga-EGFP-Nnf1b*,

respectively;  $n > 1100$ ). Both EGFP-Nnf1a and EGFP-Nnf1b are therefore functional proteins.

To compare the subcellular localization of EGFP-Nnf1a and EGFP-Nnf1b during the cell cycle, we analyzed syncytial blastoderm embryos which express maternally contributed EGFP fusion proteins. Because these embryos were derived from *Nnf1a Nnf1b* double mutant mothers rescued by either *ga-EGFP-Nnf1a* or *ga-EGFP-Nnf1b* they did not express any untagged wild-type Nnf1 protein. With both paralogs EGFP signals were clearly detectable at kinetochores throughout mitosis (Fig. 4a). In contrast, during interphase we did not observe localized EGFP signals. The same localization behavior was also observed after cellularization during progression through the embryonic cell division cycles (data not shown). While subcellular localization of EGFP-Nnf1a and EGFP-Nnf1b during the embryonic cell cycles was indistinguishable, signal intensities were different (Fig. 4b). Kinetochore signals during mitosis in case of EGFP-Nnf1b were considerably stronger than those of EGFP-Nnf1a. Independent quantifications of embryos expressing *ga-EGFP-Nnf1a* or *ga-EGFP-Nnf1b* in the presence of competing wild-type Nnf1 confirmed this difference (Fig. 4b). Immunoblotting also suggested the presence of slightly elevated levels of EGFP-Nnf1b compared to EGFP-Nnf1a (Fig. 4c) also when both were expressed from transgenes with identical regulatory sequences and chromosomal integration position. Hence Nnf1b translation efficiency and/or stability appear to be slightly higher compared to Nnf1a. In combination with the differential activity of the endogenous cis-regulatory sequences that direct production of the maternal contribution, a clear quantitative dominance of Nnf1b over Nnf1a is resulting at kinetochores during the syncytial mitoses in early embryos (Fig. 4b, c).

To evaluate the presence of the different Nnf1 proteins on meiotic kinetochores, we analyzed the EGFP tagged proteins expressed from our transgenes (*g-EGFP-Nnf1a*, *g-Nnf1a-EGFP*, *g-EGFP-Nnf1b*, and *g-Nnf1b-EGFP*) in testes and ovaries. In testes, EGFP signals

resulting from the *Nnf1b* transgenes were readily detected at kinetochores during both meiotic divisions (Fig. 5b, and data not shown). As in syncytial embryos, no localized EGFP signals were detected during interphase in spermatocytes before the onset of the meiotic divisions (Fig. 5a) and also not in spermatids after completion of the meiotic divisions (data not shown). Analogously, during progression through the gonial cell division cycles preceding meiosis, kinetochore signals were only detected during mitosis. Interestingly, a different behavior was observed for *Nnf1a*. In this case, EGFP dots were not only present during M phase (Fig. 5b) but also during interphase (Fig. 5a). Signals at centromeres during interphase were observed in spermatogonial cells (data not shown), as well as in spermatocytes before meiosis (Fig. 5a), but not in the terminally differentiated mitotically quiescent epithelial cells of the testis sheath and seminal vesicle (data not shown). During the meiotic divisions, *Nnf1a* was observed to be at kinetochores at lower levels than *Nnf1b* (Fig. 5c).

The differential localization of *Nnf1a* and *Nnf1b* during interphase was also observed in our analyses in ovaries (Fig. 5d, e). EGFP tagged *Nnf1a* was present at interphase centromeres during the gonial division cycles, in nurse cell and oocyte nuclei throughout oogenesis, as well as in the somatic follicle cells during the early stages of oogenesis as long as progression through mitotic division cycles occurs. In contrast, in case of *Nnf1b* we never observed centromeric EGFP dots during interphase. Centromeric signals during interphase were also not observed when EGFP-*Nnf1b* expression was directed by the *Nnf1a* cis-regulatory sequences (data not shown), indicating that differences in protein properties are responsible for the distinct interphase localization. However, during mitotic divisions, as well as in mature stage 14 oocytes that are arrested in metaphase of the first meiotic division, *Nnf1b* was present at kinetochores and also *Nnf1a* (Fig. 5d, e insets). We conclude that consistent with the observed genetic redundancy, *Nnf1a* and *Nnf1b* are both present at meiotic kinetochores. However, these two proteins have acquired distinct properties, at least

with regard to localization during interphase in proliferating non-embryonic cells where *Nnf1a* is centromeric but not *Nnf1b*.

### **No evidence for functional divergence of *Nnf1a* and *Nnf1b* by positive selection**

The partially redundant *D. melanogaster* *Nnf1* paralogs are distinct with regard to localization properties and expression levels during development. *Nnf1b* is particularly abundant during female meiosis. In the context of the centromere drive model it is conceivable therefore that *Nnf1b* might evolve under more prominent positive selection than *Nnf1a*. To address *Nnf1* sequence evolution, we compared the coding regions of the *Nnf1* homologs present in Drosophilid species of which whole genome sequences are known. A single homolog was found in species outside the *melanogaster* subgroup (Fig. 6) (Schittenhelm et al. 2007). In contrast, two paralogs, *Nnf1a* and *Nnf1b* were present in most of the *melanogaster* subgroup species (Fig. 6) (Schittenhelm et al. 2007). Comparison of the *Nnf1* gene phylogeny with whole genome species trees (Chen et al. 2014; Seetharam and Stuart, 2013) indicates that occurrence and fixation of the *Nnf1* gene duplication happened at the base of the branch resulting in all the sequenced species with an evolutionary origin younger than *D. kikkawai*. The paralogs *Nnf1a* and *Nnf1b* are highly diverged. The mean nucleotide distance between the two clades is 0.443 and the mean amino acid distance is 0.568. To explore if this divergence was driven by positive selection, we studied the ratio  $\omega$  of nonsynonymous substitutions  $d_N$  to synonymous substitutions  $d_S$  ( $\omega = d_N/d_S$ ) during the evolution of *Nnf1*, as reflected in the gene tree of Fig. 6. Values of this ratio above one reflect the action of positive selection. More specifically, to find out whether *Nnf1a* and *Nnf1b* are evolving under different selective pressures over the entire gene length, we first employed a class of codon substitution models known as *branch models* (Yang 1998). These models

allow the estimation of different  $\omega$  ratios for different branches in a phylogenetic tree. For this analysis we partitioned the *NnfI* phylogeny (Fig. 6) into pre-duplication branches ( $\omega_0$ ), branches immediately following the duplications ( $\omega_1$ ), branches of the clade *NnfIa* ( $\omega_2$ ), and branches of the clade *NnfIb* ( $\omega_3$ ).

We first considered the one-ratio model M0 where  $\omega$  is identical across all the branches of the *NnfI* phylogeny (Yang 2007). The estimate of  $\omega$  under this null model was 0.199, indicating that the evolution of *NnfI* was dominated by purifying selection (Table 1). A more complex model M2, where two different  $\omega$  ratios are estimated for pre-duplication ( $\omega_0$ ) and post-duplication branches ( $\omega_1 = \omega_2 = \omega_3$ ), yielded a significantly better fit than the simpler model M0 ( $2\Delta l = 14.340$ ,  $df = 1$ ,  $P = 1.53 \times 10^{-4}$ ; Table 2). This suggests that after the duplication of *NnfI*, the duplicates experienced relaxed selection that translated into an increase in the rate of evolution by a factor of 1.6 (from  $\omega = 0.145$  to  $\omega = 0.231$ ; Table 1). However, there was no evidence of positive selection acting after the duplication event, because the ratio  $\omega$  in the post-duplication branches was still below one. To explore if the rate of evolution of *NnfI* was significantly higher immediately following the gene duplication ( $\omega_1$ ) than the average rate for the other post-duplication branches ( $\omega_2 = \omega_3$ ), and the rate for the pre-duplication branches ( $\omega_0$ ), we compared a three-ratio M2 model against the two-ratio M2 model (Table 2). The estimated  $\omega$  for the  $\omega_1$  branches was 0.318 and thus higher than for the other branches (Table 1), but this more complex model did not yield a significantly better fit (Table 1;  $2\Delta l = 0.958$ ,  $df = 1$ ,  $P = 0.358$ ; Table 2). Lastly, to find out if *NnfIa* and *NnfIb* experienced different selection pressures after the gene duplication, we compared a four-ratio M2 model against the simpler two-ratio M2 model, but the differences were not significant ( $2\Delta l = 3.040$ ,  $df = 2$ ,  $P = 0.149$ ; Table 2). Together, these findings suggest that after the duplication event, the rate of evolution increased in both *NnfIa* and *NnfIb*, but the rates were not significantly different (*NnfIa* 0.251; *NnfIb* 0.205).

Although our analysis thus far had found no evidence for functional divergence of *Nnfla* and *Nnflb* through positive selection, the tests we had employed would not be able to detect signals of positive selection that occur only at one or few amino acid. To find out whether such localized signatures of selection exist, we next employed *sites models* (Nielsen and Yang 1998; Wong et al. 2004; Yang et al. 2000; Yang et al. 2005). These models allow  $\omega$  ratios to change among different codon sites, although not among different lineages. None of the models allowing sites to evolve under positive selection ( $\omega > 1$ ) showed a significantly better fit than nearly neutral models, where sites can only evolve neutrally ( $\omega = 0$ ) or experience purifying selection ( $\omega < 1$ ) (Table 3). In sum, when all the branches in the *Nnfl* phylogeny are analyzed together, there is no evidence of adaptive evolution at any particular site of this gene.

In a final analysis, we also employed *branch-site models* where  $\omega$  ratios can change both among codon sites and lineages (Yang and Nielsen 2002; Zhang et al. 2005). They allowed us to ask if positive selection could have acted on specific sites along specific branches of the *Nnfl* phylogeny (*foreground* branches). This analysis did not yield evidence for positive selection in any of the four classes of branches defined in Fig. 6 (Table 4).

In conclusion, taking together all these results, the evolution of *Nnfl* does not seem to be driven by positive selection. Yet, after the duplication event, both paralogs *Nnfla* and *Nnflb* accelerated their rates of evolution, but not significantly different.

## Discussion

Our results demonstrate that *Nnfla* and *Nnflb* are not functionally differentiated to the point where either gene can no longer provide all functions essential for development into fertile adults under laboratory conditions. However, the cis-regulatory regions of these paralogs are functionally distinct to a readily detectable degree. Expression of *Nnflb* is maximal during oogenesis and early embryogenesis when that of *Nnfla* is minimal. In contrast, *Nnfla* expression is dominant after cellularization during the embryonic cell proliferation in somatic cell lineages. Without Nnfla protein, development into fertile adults is still possible but only when two functional *Nnflb* gene copies are present. Such *Nnfla* mutant adults with two functional *Nnflb* gene copies are weak. They have also a reduced life span. These disadvantageous phenotypic consequences of a loss of *Nnfla* function are likely sufficient to explain why *Nnfla* is maintained in current *Drosophila melanogaster* populations. In contrast, our phenotypic analyses cannot explain the maintenance of *Nnflb*. Loss of *Nnflb* function does not have a readily detectable effect on viability and fertility in the laboratory. Moreover, we have failed to detect an increase in chromosome nondisjunction when analyzing X chromosome segregation during female meiosis in *Nnflb* mutants even though Nnflb appears to be predominant at kinetochores during the meiotic divisions. We emphasize that X-ND tests are far more sensitive than fertility tests. In principle, readily detectable highly significant increases in X-ND can occur in the absence of noticeable effects on fertility.

Apart from the cis-regulatory regions, the paralogous coding sequences are different as well with consequences for intracellular localization. We find that only Nnfla but not Nnflb is centromeric during interphase in gonads during production of gametes in both sexes. However, both proteins are at kinetochores during M phase. This difference between the

Nnf1 paralogs has also been observed previously after transfection of *Drosophila* cells (Przewloka et al. 2007; Schittenhelm et al. 2007). Interestingly, we find that this intracellular localization difference during interphase is developmentally controlled. During the early embryonic division cycles, Nnf1a and Nnf1b are still localized indistinguishably also during interphase where neither is centromeric.

The striking difference in the ability of Nnf1a and Nnf1b to associate with centromeres already in interphase in certain cell types might be relevant for the previously observed developmental control of *Drosophila* Mis12 localization. In embryos, Mis12 is not or at most weakly detectable at interphase centromeres, whereas it is clearly centromeric in gonads and cultured cells already during interphase (Schittenhelm et al. 2007; Venkei et al. 2012) (unpublished data). In cultured cells, centromeric localization of Mis12 and of an unidentified Nnf1 form was shown to be co-dependent (Venkei et al. 2012). Our findings implicate Nnf1a in this co-dependent localization at interphase centromeres. When Nnf1a is present at interphase centromeres, Mis12 is present as well, but not Nnf1b. In contrast, in early embryos where Nnf1a is low and unable to associate with interphase centromeres, Mis12 is also absent.

Although Nnf1a and Nnf1b are distinct with regard to centromere localization during interphase, this obvious functional difference does not appear to be of crucial physiological significance since single mutants were definitely not affected severely. Moreover, the weakness and reduced life span associated with loss of *Nnf1a* reflects reduced Nnf1 levels rather than absence of interphase centromere localization, as these traits were not observed in double mutants rescued by a single transgene copy expressing either Nnf1a or Nnf1b under *Nnf1a* control.

Localization of the Mis12 complex to interphase centromeres has also been observed in mammalian cells (Kline et al. 2006; McAinsh et al. 2006). While centromere recruitment



of the Mis12 complex during M phase is clearly essential for kinetochore function, the role of centromeric Mis12 complex components during interphase remains to be clarified. In *Drosophila*, the mechanisms for centromere recruitment during interphase and mitosis have been demonstrated to be distinct (Venkei et al. 2012). Centromeric Mis12 complex components might act in a functionally redundant manner in heterochromatin regulation, since the heterochromatin protein HP1 has been demonstrated to recruit the Mis12 complex to interphase centromeres in mammalian cells (Kiyomitsu et al. 2010; Obuse et al. 2004). Accordingly, the absence of heterochromatin during the early syncytial cycles of *Drosophila* embryogenesis (Rudolph et al. 2007), when extremely rapid DNA replication in the absence of widespread transcription prevails, might explain the absence of Nnf1a from interphase centromeres specifically during the initial embryonic stages.

Our characterization of the *Nnf1* paralogs provides a basis for future analyses addressing the evolutionary significance of gene duplications which have attracted increasing interest ever since Ohno's seminal publication (Ohno 1970). Our comparison of *Drosophilid* sequences suggests that the *Nnf1* duplication was followed by accelerated sequence divergence. However, we have been unable to detect positive selection. Positive selection in particular in the *Nnf1b* branch would have been of great interest in the context of the centromere drive hypothesis (Henikoff et al. 2001). *Nnf1b* has properties that appear to make it a good target for mutations beneficial for suppression of preferential meiotic segregation of centromere variants into the female pronucleus. Nnf1b predominates in kinetochores during female meiosis where it forms a bridge (Przewłoka et al. 2011) between Cenp-C, an inner centromere protein with noticeable direct DNA binding activity (Hori et al. 2008; Politi et al. 2002; Sugimoto et al. 1994; Yang et al. 1996), and the rest of the KMN network, the main microtubule-binding activity that might have a prominent role in the assembly and positioning of the acentrosomal spindles in oocytes (Radford et al. 2015). Clearly, the

absence of evident positive selection in the *Nnf1b* branch does not argue against centromere drive as an important evolutionary force within the *D. melanogaster* lineage where positive selection in the *cid/Cenp-A* gene has actually been reported in support of the centromere drive hypothesis (Cooper and Henikoff 2004; Malik and Henikoff 2001). Interestingly, a *CENH3/Cenp-A* duplication has been recently described in *Mimulus*, where centromere drive is clearly ongoing, and differential evolution of the paralogs has been detected in this case (Finseth et al. 2015). Transient positive selection in the context of centromere drive early after *Nnf1* duplication cannot be ruled out by our data and might therefore still have been important for the retention of the duplicated *Nnf1* genes. Apart from centromere drive, an adaptive conflict caused by distinct selective pressures on Nnf1 function during meiosis and mitosis, for example, might have favored retention of the duplicated copies by permitting functional optimization of Nnf1b for meiosis. Nnf1b specialization for the extremely rapid mitoses of early embryogenesis might be another conceivable alternative. If still present in *D. melanogaster*, such Nnf1b specialization would have to be subtle, as we have not detected obvious effects after loss of *Nnf1b*. However, since Nnf1b is predominant during both female and male meiosis, we favor meiotic specialization apart from the evident partial subfunctionalization of the cis-regulatory regions as probable explanations for the maintenance of the functionally overlapping Nnf1 paralogs.

## Acknowledgements

We thank Frank Sprenger for supporting M.K. during the initial work, and Martina Trost for technical help. This work was supported by a Swiss National Science Foundation grant (31003A\_120276) to C.F.L., a Swiss National Science Foundation grant (31003A\_146137) to

A.W., a grant from the Forschungskredit program of the University of Zurich (FK-14-076) to J.A.-R., and A.W.'s support by the University Priority Research Program in Evolutionary Biology at the University of Zurich.

## **Compliance with Ethical Standards**

This article does not contain any studies performed by any of the authors with human participants or animals (except for the invertebrate *Drosophila melanogaster* which is not subject to animal research legislation).

## **Conflict of Interest**

The authors declare that they have no competing interests.

## References

- Abbey HN, Kral LG (2015) Adaptive Evolution of CENP-A in Percid Fishes. *Genes* 6:662-671. doi:10.3390/genes6030662
- Akiyoshi B, Gull K (2014) Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156:1247-1258. doi:10.1016/j.cell.2014.01.049
- Axelsson E et al. (2010) Segregation distortion in chicken and the evolutionary consequences of female meiotic drive in birds. *Heredity* 105:290-298. doi:10.1038/hdy.2009.193
- Beck EA, Llopart A (2015) Widespread Positive Selection Drives Differentiation of Centromeric Proteins in the *Drosophila melanogaster* subgroup. *Scientific reports* 5:17197. doi:10.1038/srep17197
- Bellen HJ et al. (2004) The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* 167:761-781. doi:10.1534/genetics.104.026427
- Berriman M et al. (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309:416-422. doi:10.1126/science.1112642
- Bischof J, Bjorklund M, Furger E, Schertel C, Taipale J, Basler K (2013) A versatile platform for creating a comprehensive UAS-ORFeome library in *Drosophila*. *Development* 140:2434-2442. doi:10.1242/dev.088757
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-1973. doi:10.1093/bioinformatics/btp348
- Chen ZX et al. (2014) Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res* 24:1209-1223. doi:10.1101/gr.159384.113

- Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* 39:715-720. doi:10.1038/ng2049
- Chmatal L et al. (2014) Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr Biol* 24:2295-2300. doi:10.1016/j.cub.2014.08.017
- Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. *Molecular biology and evolution* 21:1712-1718
- Drinnenberg IA, deYoung D, Henikoff S, Malik HS (2014) Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife* 3. doi:10.7554/eLife.03676
- Ejsmont RK, Sarov M, Winkler S, Lipinski KA, Tomancak P (2009) A toolkit for high-throughput, cross-species gene engineering in *Drosophila*. *Nat Methods* 6:435-437. doi:10.1038/nmeth.1334
- Fachinetti D et al. (2013) A two-step mechanism for epigenetic specification of centromere identity and function. *Nat Cell Biol* 15:1056-1066. doi:10.1038/ncb2805
- Finseth FR, Dong Y, Saunders A, Fishman L (2015) Duplication and Adaptive Evolution of a Key Centromeric Protein in *Mimulus*, a Genus with Female Meiotic Drive. *Molecular biology and evolution*. doi:10.1093/molbev/msv145
- Fishman L, Saunders A (2008) Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* 322:1559-1562. doi:10.1126/science.1161406
- Fukagawa T, Earnshaw WC (2014) The centromere: chromatin foundation for the kinetochore machinery. *Dev Cell* 30:496-508. doi:10.1016/j.devcel.2014.08.016
- Gascoigne KE, Takeuchi K, Suzuki A, Hori T, Fukagawa T, Cheeseman IM (2011) Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* 145:410-422. doi:10.1016/j.cell.2011.03.031

- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular biology and evolution* 11:725-736
- Graveley BR et al. (2011) The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471:473-479. doi:10.1038/nature09715
- Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* 166:1775-1782
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59:307-321.  
doi:10.1093/sysbio/syq010
- Guse A, Carroll CW, Moree B, Fuller CJ, Straight AF (2011) In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* 477:354-358.  
doi:10.1038/nature10379
- Heeger S, Leismann O, Schittenhelm R, Schraidt O, Heidmann S, Lehner CF (2005) Genetic interactions of Separase regulatory subunits reveal the diverged *Drosophila* Cenp-C homolog. *Genes Dev* 19:2041-2053
- Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293:1098-1102.
- Henikoff S, Ahmad K, Platero JS, van Steensel B (2000) Heterochromatic deposition of centromeric histone H3-like proteins. *Proc Natl Acad Sci U S A* 97:716-721
- Hori T et al. (2008) CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* 135:1039-1052.  
doi:10.1016/j.cell.2008.10.019

- Hori T, Shang WH, Takeuchi K, Fukagawa T (2013) The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *J Cell Biol* 200:45-60. doi:10.1083/jcb.201210106
- Horn C, Jaunich B, Wimmer EA (2000) Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev Genes Evol* 210:623-629
- Hornung P et al. (2014) A cooperative mechanism drives budding yeast kinetochore assembly downstream of CENP-A. *J Cell Biol* 206:509-524. doi:10.1083/jcb.201403081
- Hoskins RA et al. (2000) A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science* 287:2271-2274
- Huelsenbeck JP, Crandall KA (1997) Phylogeny Estimation and Hypothesis Testing using Maximum Likelihood. *Annu Rev Ecol Syst* 28:437-466
- Jacobs HW, Knoblich JA, Lehner CF (1998) *Drosophila* Cyclin B3 is required for female fertility and is dispensable for mitosis like Cyclin B. *Genes Dev* 12:3741-3751
- Karpen GH, Allshire RC (1997) The case for epigenetic effects on centromere identity and function. *Trends Genet* 13:489-496. doi:S0168952597012985 [pii]
- Kim S, Yu H (2015) Multiple assembly mechanisms anchor the KMN spindle checkpoint platform at human mitotic kinetochores. *J Cell Biol* 208:181-196. doi:10.1083/jcb.201407074
- Kiyomitsu T, Iwasaki O, Obuse C, Yanagida M (2010) Inner centromere formation requires hMis14, a trident kinetochore protein that specifically recruits HP1 to human chromosomes. *J Cell Biol* 188:791-807. doi:10.1083/jcb.200908096
- Klare K et al. (2015) CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J Cell Biol* 210:11-22. doi:10.1083/jcb.201412028

- Kline SL, Cheeseman IM, Hori T, Fukagawa T, Desai A (2006) The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. *J Cell Biol* 173:9-17
- Le T et al. (2006) A new family of *Drosophila* balancer chromosomes with a w- dfd-GMR yellow fluorescent protein marker. *Genetics* 174:2255-2257.  
doi:10.1534/genetics.106.063461
- London N, Ceto S, Ranish JA, Biggins S (2012) Phosphoregulation of Spc105 by Mps1 and PP1 regulates Bub1 localization to kinetochores. *Curr Biol* 22:900-906.  
doi:10.1016/j.cub.2012.03.052
- Löytynoja A, Goldman N (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320:1632-1635.  
doi:10.1126/science.1158395
- Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* 157:1293-1298
- Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138:1067-1082. doi:10.1016/j.cell.2009.08.036
- Malvezzi F, Litos G, Schleiffer A, Heuck A, Mechtler K, Clausen T, Westermann S (2013) A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. *EMBO J* 32:409-423. doi:10.1038/emboj.2012.356
- Marshall OJ, Chueh AC, Wong LH, Choo KH (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* 82:261-282. doi:10.1016/j.ajhg.2007.11.009
- McAinsh AD, Meraldi P, Draviam VM, Toso A, Sorger PK (2006) The human kinetochore proteins Nnf1R and Mcm21R are required for accurate chromosome segregation. *Embo J* 25:4033-4049



- Mendiburo MJ, Padeken J, Fulop S, Schepers A, Heun P (2011) *Drosophila* CENH3 is sufficient for centromere formation. *Science* 334:686-690.  
doi:10.1126/science.1206880
- Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol* 7:R23
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929-936
- Nishino T, Rago F, Hori T, Tomii K, Cheeseman IM, Fukagawa T (2013) CENP-T provides a structural platform for outer kinetochore assembly. *EMBO J* 32:424-436.  
doi:10.1038/emboj.2012.348
- Obuse C, Iwasaki O, Kiyomitsu T, Goshima G, Toyoda Y, Yanagida M (2004) A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nat Cell Biol* 6:1135-1141
- Ohno S (1970) *Evolution by gene duplication*. Springer-Verlag, New York
- Parks AL et al. (2004) Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nat Genet* 36:288-292. doi:10.1038/ng1312
- Petrovic A et al. (2014) Modular assembly of RWD domains on the Mis12 complex underlies outer kinetochore organization. *Mol Cell* 53:591-605.  
doi:10.1016/j.molcel.2014.01.019
- Petrovic A et al. (2010) The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *J Cell Biol* 190:835-852. doi:10.1083/jcb.201002070
- Politi V, Perini G, Trazzi S, Pliss A, Raska I, Earnshaw WC, Della Valle G (2002) CENP-C binds the alpha-satellite DNA in vivo at specific centromere domains. *J Cell Sci* 115:2317-2327

- Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular biology and evolution* 25:1253-1256. doi:10.1093/molbev/msn083
- Primorac I et al. (2013) Bub3 reads phosphorylated MELT repeats to promote spindle assembly checkpoint signaling. *Elife* 2:e01030. doi:10.7554/eLife.01030
- Przewłoka MR, Glover DM (2009) The kinetochore and the centromere: a working long distance relationship. *Annu Rev Genet* 43:439-465. doi:10.1146/annurev-genet-102108-134310
- Przewłoka MR, Venkei Z, Bolanos-Garcia VM, Debski J, Dadlez M, Glover DM (2011) CENP-C is a structural platform for kinetochore assembly. *Curr Biol* 21:399-405. doi:10.1016/j.cub.2011.02.005
- Przewłoka MR et al. (2007) Molecular analysis of core kinetochore composition and assembly in *Drosophila melanogaster*. *PLoS ONE* 2:e478. doi:10.1371/journal.pone.0000478
- Radermacher PT, Myachina F, Bosshardt F, Pandey R, Mariappa D, Muller HA, Lehner CF (2014) O-GlcNAc reports ambient temperature and confers heat resistance on ectotherm development. *Proc Natl Acad Sci U S A* 111:5592-5597. doi:10.1073/pnas.1322396111
- Radford SJ, Hoang TL, Gluszek AA, Ohkura H, McKim KS (2015) Lateral and End-On Kinetochore Attachments Are Coordinated to Achieve Bi-orientation in *Drosophila* Oocytes. *PLoS Genet* 11:e1005605. doi:10.1371/journal.pgen.1005605
- Rago F, Gascoigne KE, Cheeseman IM (2015) Distinct organization and regulation of the outer kinetochore KMN network downstream of CENP-C and CENP-T. *Curr Biol* 25:671-677. doi:10.1016/j.cub.2015.01.059

- Rudolph T et al. (2007) Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell* 26:103-115. doi:10.1016/j.molcel.2007.02.025
- Schittenhelm RB, Althoff F, Heidmann S, Lehner CF (2010) Detrimental incorporation of excess Cenp-A/Cid and Cenp-C into *Drosophila* centromeres is prevented by limiting amounts of the bridging factor Cal1. *J Cell Sci* 123:3768-3779. doi:10.1242/jcs.067934
- Schittenhelm RB, Chaleckis R, Lehner CF (2009) Essential functional domains and intrakinetochore localization of *Drosophila* Spc105. *EMBO J* 28:2374-2386
- Schittenhelm RB, Heeger S, Althoff F, Walter A, Heidmann S, Mechtler K, Lehner CF (2007) Spatial organization of a ubiquitous eukaryotic kinetochore protein network in *Drosophila* chromosomes. *Chromosoma* 116:385-402
- Schleiffer A, Maier M, Litos G, Lampert F, Hornung P, Mechtler K, Westermann S (2012) CENP-T proteins are conserved centromere receptors of the Ndc80 complex. *Nat Cell Biol* 14:604-613. doi:10.1038/ncb2493
- Schueler MG, Swanson W, Thomas PJ, Program NCS, Green ED (2010) Adaptive evolution of foundation kinetochore proteins in primates. *Molecular biology and evolution* 27:1585-1597. doi:10.1093/molbev/msq043
- Schuldiner O, Berdnik D, Levy JM, Wu JS, Luginbuhl D, Gontang AC, Luo L (2008) piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev Cell* 14:227-238. doi:10.1016/j.devcel.2007.11.001
- Screpanti E, De Antoni A, Alushin GM, Petrovic A, Melis T, Nogales E, Musacchio A (2011) Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol* 21:391-398. doi:10.1016/j.cub.2010.12.039

- Seetharam AS, Stuart GW (2013) Whole genome phylogeny for 21 *Drosophila* species using predicted 2b-RAD fragments. *PeerJ* 1:e226. doi:DOI 10.7717/peerj.226
- Shan X, Xue Z, Euskirchen G, Melese T (1997) NNF1 is an essential yeast gene required for proper spindle orientation, nucleolar and nuclear envelope structure and mRNA export. *J Cell Sci* 110 ( Pt 14):1615-1624
- Shepherd LA et al. (2012) Phosphodependent recruitment of Bub1 and Bub3 to Spc7/KNL1 by Mph1 kinase maintains the spindle checkpoint. *Curr Biol* 22:891-899. doi:10.1016/j.cub.2012.03.051
- Sugimoto K, Yata H, Muro Y, Himeno M (1994) Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. *J Biochem (Tokyo)* 116:877-881
- Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol* 3:18
- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14:1053-1066
- Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* 101:11030-11035. doi:10.1073/pnas.0404206101
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30:2725-2729. doi:10.1093/molbev/mst197
- Thibault ST et al. (2004) A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. *Nat Genet* 36:283-287
- Venkei Z et al. (2012) Spatiotemporal dynamics of Spc105 regulates the assembly of the *Drosophila* kinetochore. *Open biology* 2:110032. doi:10.1098/rsob.110032

- Venken KJ, He Y, Hoskins RA, Bellen HJ (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* 314:1747-1751. doi:10.1126/science.1134426
- Voullaire LE, Slater HR, Petrovic V, Choo KH (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* 52:1153-1163
- Westermann S, Cheeseman IM, Anderson S, Yates JR, 3rd, Drubin DG, Barnes G (2003) Architecture of the budding yeast kinetochore reveals a conserved molecular core. *J Cell Biol* 163:215-222
- Westermann S, Schleiffer A (2013) Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends Cell Biol* 23:260-269. doi:10.1016/j.tcb.2013.01.010
- Westhorpe FG, Straight AF (2015) The centromere: epigenetic control of chromosome segregation during mitosis. *Cold Spring Harbor perspectives in biology* 7:a015818. doi:10.1101/cshperspect.a015818
- White-Cooper H (2004) Spermatogenesis: analysis of meiosis and morphogenesis. *Methods Mol Biol* 247:45-75. doi:1-59259-665-7-45 [pii]
- Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041-1051. doi:10.1534/genetics.104.031153
- Yang CH, Tomkiel J, Saitoh H, Johnson DH, Earnshaw WC (1996) Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. *Mol Cell Biol* 16:3576-3586
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and evolution* 15:568-573

- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24:1586-1591. doi:10.1093/molbev/msm088
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular biology and evolution* 19:908-917
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular biology and evolution* 22:1107-1118. doi:10.1093/molbev/msi097
- Yuan J, Guo X, Hu J, Lv Z, Han F (2015) Characterization of two CENH3 genes and their roles in wheat evolution. *The New phytologist* 206:839-851. doi:10.1111/nph.13235
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular biology and evolution* 22:2472-2479. doi:10.1093/molbev/msi237
- Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) *Evolving Genes and Proteins*. Academic Press, New York, pp 97–166

## Figure legends

**Fig. 1** *Nnf1b* is not essential for viability and fertility.

**a** Scheme of the *Nnf1b* region in the *Drosophila melanogaster* genome. *Nnf1b* and the three downstream genes *Dbp21E2*, *Saf6* and *Pex12* are shown (open boxes: untranslated regions, filled boxes coding regions), as well as insertion positions of transposons (triangles) used for the generation of the deficiencies *Df(2L)MK01* and *Df(2L)NK01* (black bars). The genomic region present in the *gDSP* transgene is indicated as well (open bar).

**b** A multiplex PCR assay confirmed absence of *Nnf1b* in *Nnf1b<sup>null-2</sup>* mutant flies. Products amplified from *w* control (+) or *Nnf1b<sup>null-2</sup>* (*null-2*) genomic DNA, or in the absence of template DNA (H<sub>2</sub>O) were analyzed by agarose gel electrophoresis. A molecular weight marker (M) was used for size comparison of the fragments amplified from *Nnf1a* and *Nnf1b* (arrowheads).

**c** Loss of *Nnf1b* does not interfere with development to the adult stage. Homozygous *Df(2L)MK01*, *gDSP II.2* females were crossed with males that were either *Df(2L)NK01/CyO* (*null-1*) or *Df(2L)NK01*, *gDSP II.2/CyO* (*null-2*) and the number of adult progeny with (+) and without (-) CyO was determined and is displayed (% of total adult progeny).

**d** Fertility in the absence of *Nnf1b* is normal. The fertility of males and females with the genotypes *w* for control (+), *Nnf1b<sup>null-1</sup>* (*null-1*) or *Nnf1b<sup>null-2</sup>* (*null-2*) was determined. Fertility of controls was set to 100%. Relative average (bars) and standard deviation (whiskers) are displayed (n = 10).

**e** Loss of *Nnf1b* does not increase the rate of X chromosome nondisjunction (X-ND) during female meiosis. The rate of X-ND (in % of total meioses) in females with the genotypes *w* for control (+), *Nnf1b<sup>null-1</sup>* (*null-1*) or *Nnf1b<sup>null-2</sup>* (*null-2*) was determined (see Materials and Methods). Based on distinct adult visible phenotypes, males (m regular) and females (f

regular) resulting from euploid oocytes as well as adult progeny (irregular) indicating X-ND were scored and counted.

**Fig. 2** *Nnf1a* is not essential for viability and fertility.

**a** Scheme illustrating the structure of the *Nnf1a* gene (open boxes: untranslated regions, filled boxes: coding regions) as well as the insertion position of the *PBacLL02791* transposon (triangle) which includes a splice acceptor site (SA) followed by stop codons (stop) in all three reading frames. Primer pairs used for the analysis of the effects of the transposon insertion on the *Nnf1a* transcript splice pattern are indicated by colored arrows.

**b** The *PBacLL02791* insertion prevents production of normally spliced *Nnf1a* transcripts. RNA isolated from embryos during the syncytial stages was analyzed by qRT-PCR with the primer pairs illustrated in (a) and an additional *Nnf1b*-specific primer pair (magenta). The isolated RNA is of maternal origin. The mothers had the genotype *w* for control (+), *PBacLL02791/CyO* (*PBac/CyO*), *PBacLL02791* (*PBac*) or *Df(2L)MK01/Df(2L)NK01; gDSP III.2/+* (*Nnf1b* null). Product amount detected in *PBac/CyO* with a given primer pair was set to 100%. Relative average (bars) and standard deviation (whiskers) are displayed ( $n \geq 3$ ). (n.d.) not determined.

**c** Loss of *Nnf1a* does not interfere with development to the adult stage. *PBacLL02791/CyO* virgin females were crossed with males that were either *PBacLL02791/CyO* (*PBac*), *Df(2R)Exel6070/CyO* (*Df6070*) or *Df(2R)Exel7164/CyO* (*Df7164*) and the number of adult progeny with (+) and without (-) CyO was determined and is displayed (% of total adult progeny).

**d** Fertility is not affected severely by absence of *Nnf1a*. The fertility of males and females with the genotypes *w* for control (+), *PBacLL02791* (*PBac*), *PBacLL02791/Df(2R)Exel6070* (*Df6070*), or *PBacLL02791/Df(2R)Exel7164* (*Df7164*) was determined. Fertility of controls



was set to 100%. Relative average (bars) and standard deviation (whiskers) are displayed (n = 10).

**Fig. 3** Simultaneous loss of *Nnf1a* and *Nnf1b* results in embryonic abnormalities. Embryos were collected from a cross of *Df(2L)MK01*, *gDSP II.2*, *PBacLL02791/CyO*, *Dfd-EYFP* females with *Df(2L)NK01*, *gDSP II.2*, *Df(2R)Exel7164/CyO*, *Dfd-EYFP* males. Embryos were fixed and stained with anti-Cyclin B (CycB) and a DNA stain (DNA). *Nnf1a Nnf1b* double mutants (*Nnf1a<sup>-</sup> Nnf1b<sup>-</sup>*), which did not express Dfd-EYFP in the head region unlike balanced sibling embryos (*Nnf1a<sup>+</sup> Nnf1b<sup>+</sup>*) (arrowhead), contained some cells in the CNS that were enlarged and/or expressed Cyclin B at higher levels. Bottom row shows single sections from CNS regions indicated in the top row (dashed rectangles) at higher magnification. Scale bars correspond to 100  $\mu$ m (top) and 20  $\mu$ m (bottom), respectively.

**Fig. 4** *Nnf1b* accumulates to higher levels than *Nnf1a* at kinetochores of syncytial embryos.

**a** During mitosis but not during interphase, EGFP-*Nnf1a* and EGFP-*Nnf1b* both localize to kinetochores in syncytial blastoderm embryos generated by *Nnf1a Nnf1b* double mutant females carrying two transgene copies of either *ga-EGFP-Nnf1a* (upper part) or *ga-EGFP-Nnf1b* (lower part). The metaphase plates indicated by the arrowheads are shown at higher magnification in the insets. Scale bar: 10  $\mu$ m.

**b** The intensity of EGFP signals at kinetochores in syncytial blastoderm embryos during prometa- and metaphase were quantified. The genotypes of the mothers are indicated below the bars. Two copies of *ga-EGFP-Nnf1a* (*ga-E-Nnf1a*) or *ga-EGFP-Nnf1b* (*ga-E-Nnf1b*) were present in either a *Nnf1a Nnf1b* double mutant background (*null*) or in a background with *Nnf1a* and *Nnf1b* function (+). In the latter background, we also analyzed single copies

of *g-EGFP-Nnfla* (*g-E-Nnfla*), *g-Nnfla-EGFP* (*g-Nnfla-E*), *g-EGFP-Nnflb* (*g-E-Nnflb*), or *g-Nnflb-EGFP* (*g-Nnflb-E*).

**c** EGFP-Nnflb levels are higher than those of EGFP-Nnfla also when expressed under control of the same cis-regulatory sequences. Total extracts from syncytial blastoderm embryos were analyzed by immunoblotting with anti-EGFP and anti-PSTAIR used as loading control. The embryos were collected from mothers carrying either *ga-EGFP-Nnfla* (*ga-E-Nnfla*), *ga-EGFP-Nnflb* (*ga-E-Nnflb*), *g-EGFP-Nnfla* (*g-E-Nnfla*), or *g-EGFP-Nnflb* (*g-E-Nnflb*), or no transgene (+). All transgenes were homozygous except *g-EGFP-Nnflb* which was present in one copy. For accurate quantification, a dilution series of the *ga-EGFP-Nnflb* extract was loaded in the first three lanes. Bar diagram represent expression levels (mean and s. d.) resulting from immunoblot quantification (n = 3) where levels of the different EGFP fusion proteins were quantified relative to those obtained with *ga-EGFP-Nnflb*. EGFP-Nnflb levels were set to 100 in each blot.

**Fig. 5** Nnfla but not Nnflb accumulates at centromeres during interphase in testes and ovaries.

**a-c** Images of mature stage S6 spermatocytes before entry into the meiotic divisions (**a**) and of prometaphase cells of the first meiotic division (**b**) were acquired from squash preparations of *g-EGFP-Nnfla* and *g-EGFP-Nnflb* testes. EGFP-Nnfla but not EGFP-Nnflb was detected at centromeres in S6 spermatocytes (**a**). However, both were detected at kinetochores during the meiotic divisions (**b**), and quantification of meiotic signals (**c**) confirmed that EGFP-Nnflb levels were higher than those of EGFP-Nnfla. Bars represent average, whiskers s.d. n > 43 spermatocytes in prometaphase or metaphase of meiosis I.

**d, e** Images of germaria (**d**) and egg chambers at stage S4 (**e**) and S14 (insets in **e**) were acquired from *g-EGFP-Nnfla* and *g-EGFP-Nnflb* ovarioles. While EGFP-Nnfla was

detected at centromeres in all germline cells within the germarium, EGFP-Nnf1b was only detected at kinetochores in cysts progressing through mitosis. Similarly, at later stages, EGFP-Nnf1a was centromeric in follicle cells (arrowhead f), nurse cells (arrowhead n) and oocyte nucleus (arrowhead o), while EGFP-Nnf1b was not detectable at centromeres. However, in mature stage S14 oocytes which are arrested at metaphase of the first meiotic division, EGFP-Nnf1a and EGFP-Nnf1b were both detected at kinetochores. All scale bars: 10  $\mu$ m. Metaphase I figures from S14 oocytes in the insets (e) are shown with 1.7 fold higher magnification.

**Fig. 6.** Maximum-likelihood phylogenetic reconstruction for the gene *Nnf1*. The gene tree was obtained using PhyML v3 (Guindon et al. 2010) under the nucleotide substitution model “TIM1 + I +  $\Gamma$ ”, using 1000 bootstrap replicates. The root of the tree is based on a recent phylogeny for *Drosophila* species (Seetharam and Stuart 2013). The asterisk indicates the duplication of *Nnf1*. The clades corresponding to the paralogs *Nnf1a* and *Nnf1b* are highlighted in orange and yellow shading, respectively. For the branch and branch-site codon substitution models, we subdivided the branches in the gene tree into pre-duplication branches ( $\omega_0$ , solid black), branches immediately following the duplication ( $\omega_1$ , gray), branches of the clade *Nnf1a* ( $\omega_2$ , dashed), and branches of the clade *Nnf1b* ( $\omega_3$ , bold). Red circles at nodes indicate bootstrap support values greater than 90%. The scale bar represents nucleotide substitutions per site.

**Table 1.** Parameter estimates and log-likelihood values under different branch models for the evolution of Nnfl

<b>Model<sup>a)</sup></b>	<b>Estimated parameters<sup>b)</sup></b>	<b>Log-likelihood (l)</b>
M0: One-ratio	$\omega_0 = \omega_1 = \omega_2 = \omega_3 = 0.19873$	-8056.613025
M2: Two-ratio	$\omega_0 = 0.14473, \omega_1 = \omega_2 = \omega_3 = 0.23133$	-8049.443005
M2: Three-ratio	$\omega_0 = 0.14456, \omega_1 = 0.31844, \omega_2 = \omega_3 = 0.22696$	-8048.963903
M2: Four-ratio	$\omega_0 = 0.14461, \omega_1 = 0.32065, \omega_2 = 0.25121, \omega_3 = 0.20458$	-8047.922804

a) Models from Yang (1998)

b) See branch-specific  $\omega$  ratios in Fig. 6

**Table 2. Likelihood ratio tests of branch models**

<b>Model comparisons<sup>a)</sup></b>	<b><math>2\Delta I^{b)}</math></b>	<b>df<sup>c)</sup></b>
M0 (One-ratio) vs. M2 (Two-ratio)	14.340***	1
M2 (Two-ratio) vs. M2 (Three-ratio)	3.040	1
M2 (Two-ratio) vs. M2 (Four-ratio)	2.081	2

a) Models from Yang (1998)

b) Significance: \*\*\* $P < 0.001$

c) Degrees of freedom

**Table 3. Likelihood ratio tests of sites models**

<b>Model comparisons<sup>a)</sup></b>	<b>2Δl<sup>b)</sup></b>	<b>df<sup>c)</sup></b>
M1a vs. M2a	$-3.38 \times 10^{-3}$	2
M7 vs. M8	$-6.60 \times 10^{-5}$	2

a) See Material and Methods for details

b) None of these model comparisons indicated a significant difference

c) Degrees of freedom

**Table 4. Likelihood ratio tests of branch-site models**

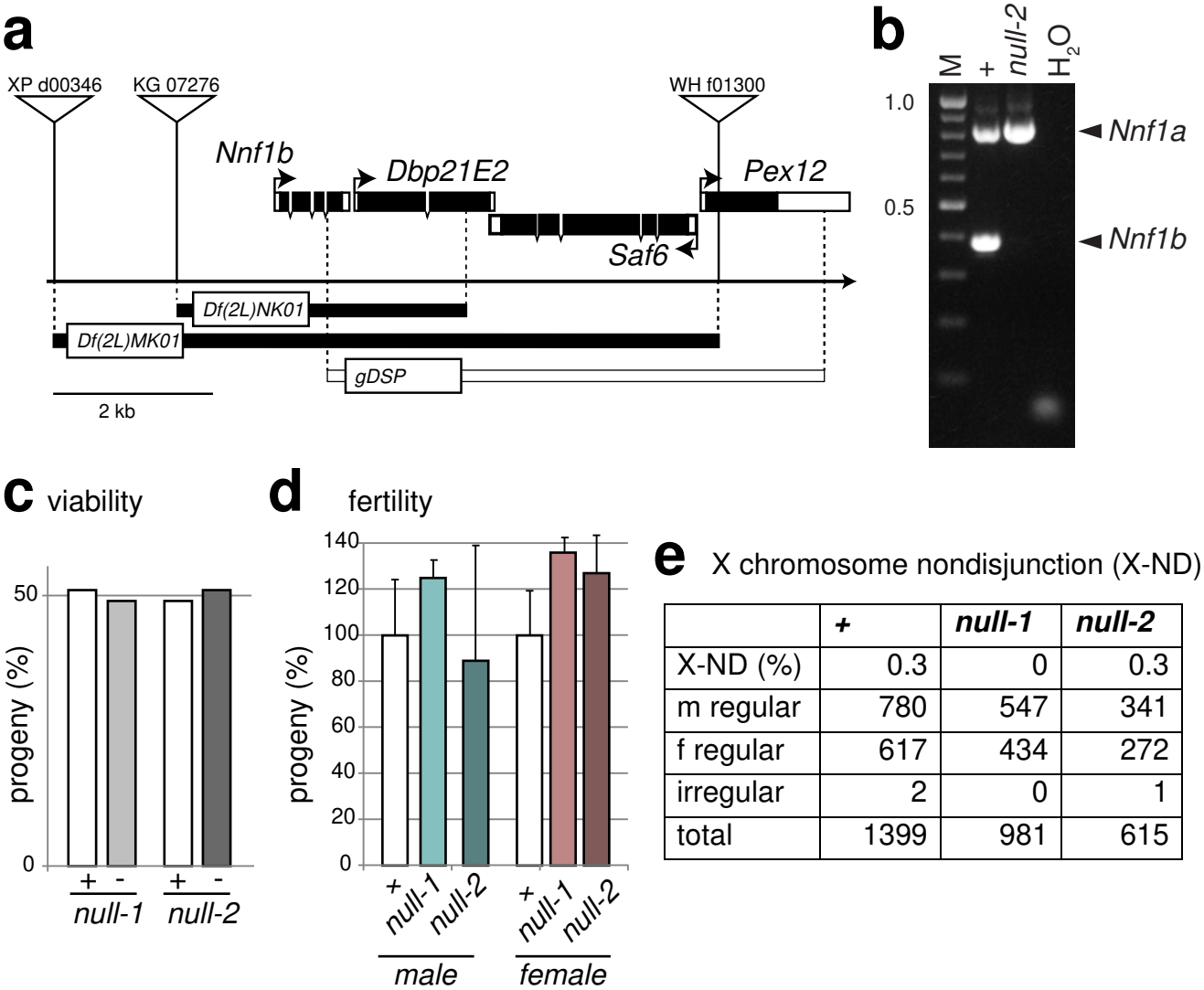
<b>Foreground branch<sup>a)</sup></b>	<b><math>2\Delta l</math> (A vs. <math>A_{\text{null}}</math>)<sup>b,c)</sup></b>	<b>df<sup>d)</sup></b>
$\omega_0$	2.723	1
$\omega_1$	0.000	1
$\omega_2$	0.000	1
$\omega_3$	0.000	1

a) See branch names in Fig. 6

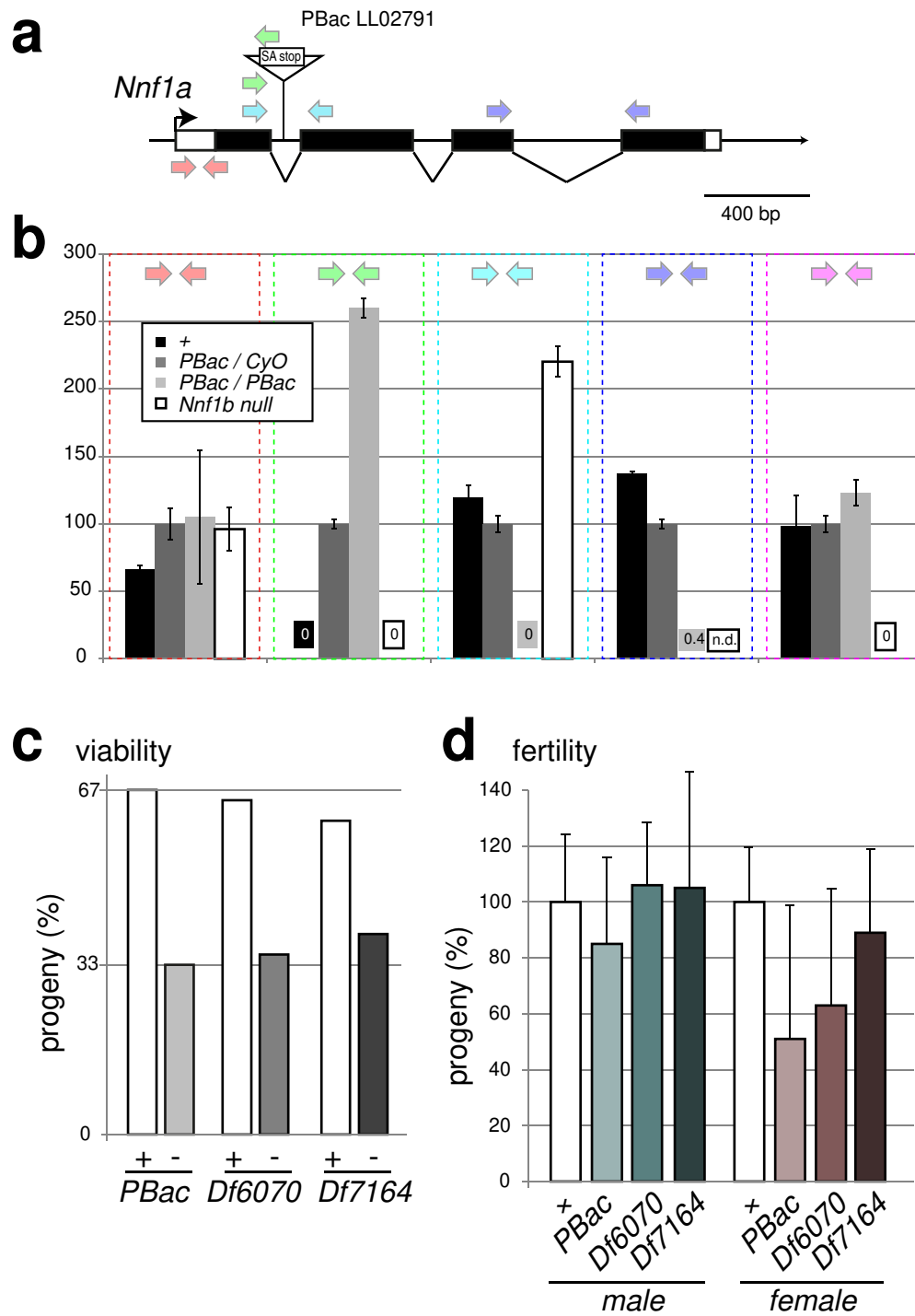
b) Models from Zhang et al. (2005)

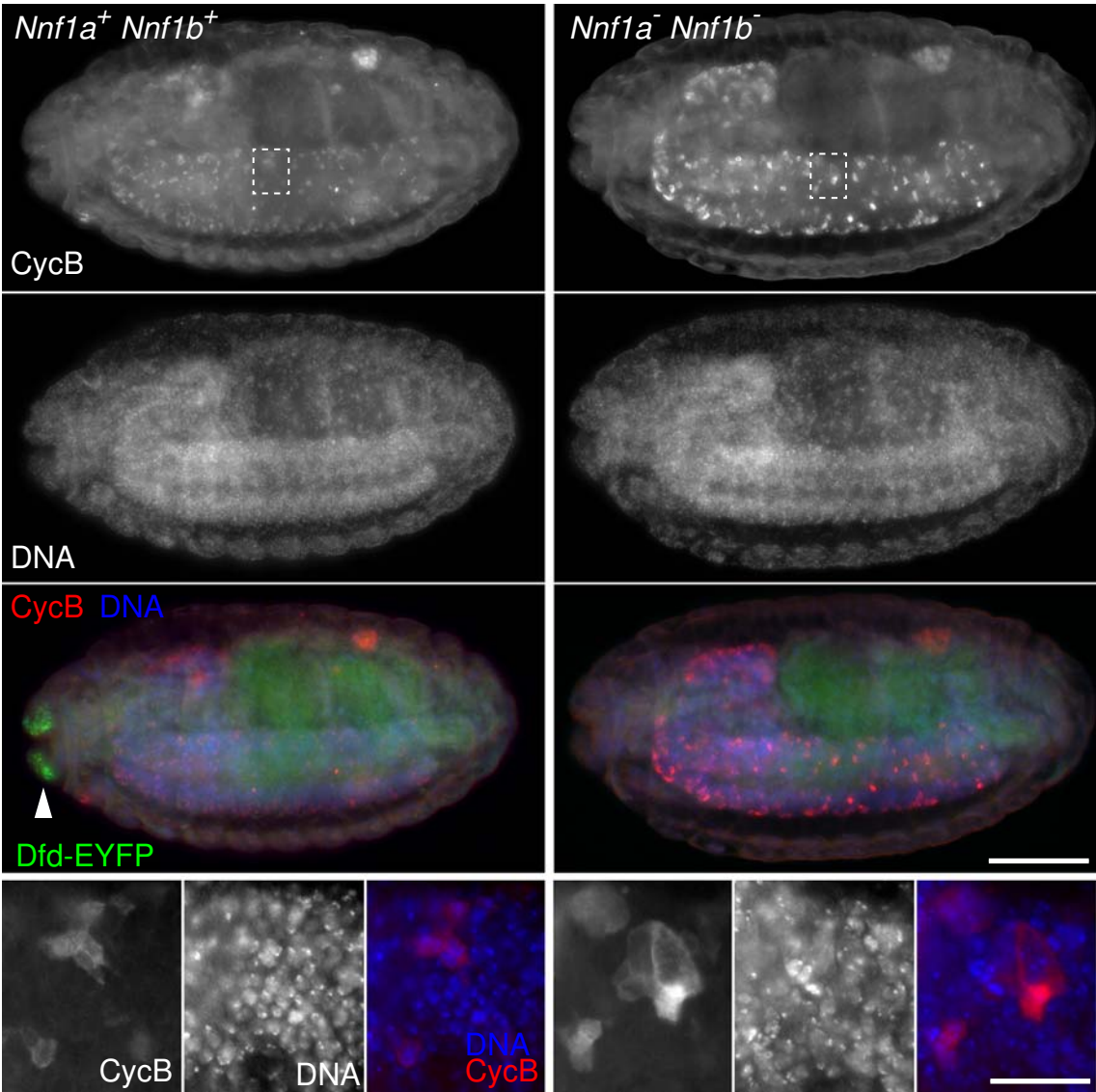
c) None of these model comparisons indicated a significant difference

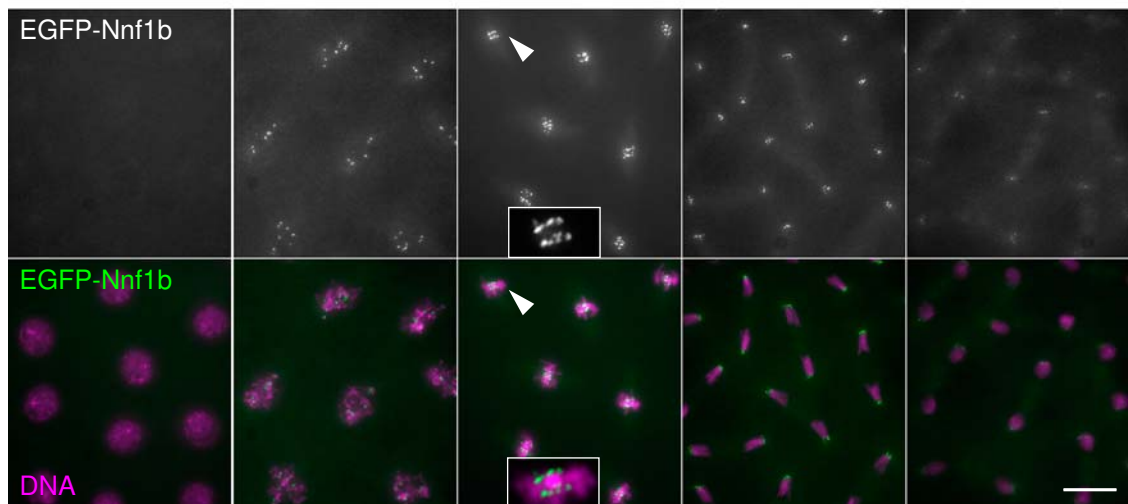
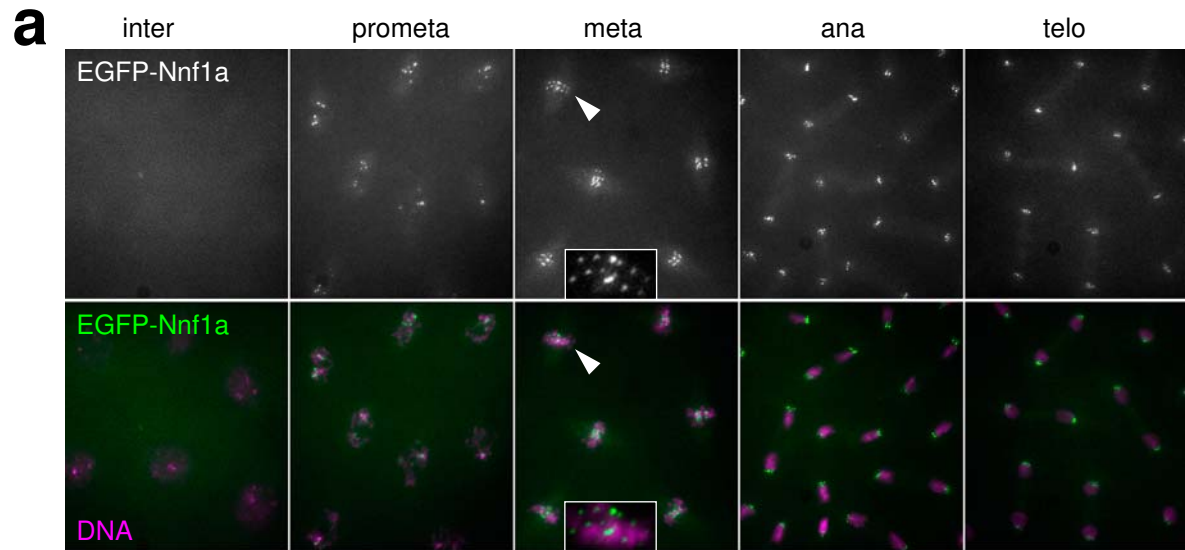
d) Degrees of freedom



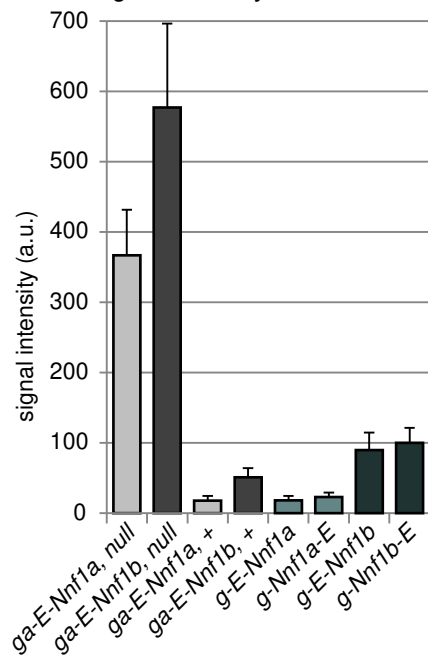




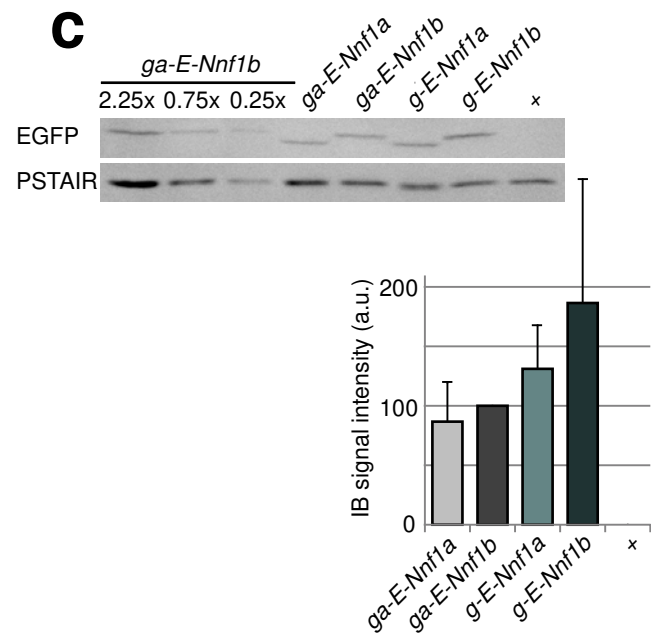


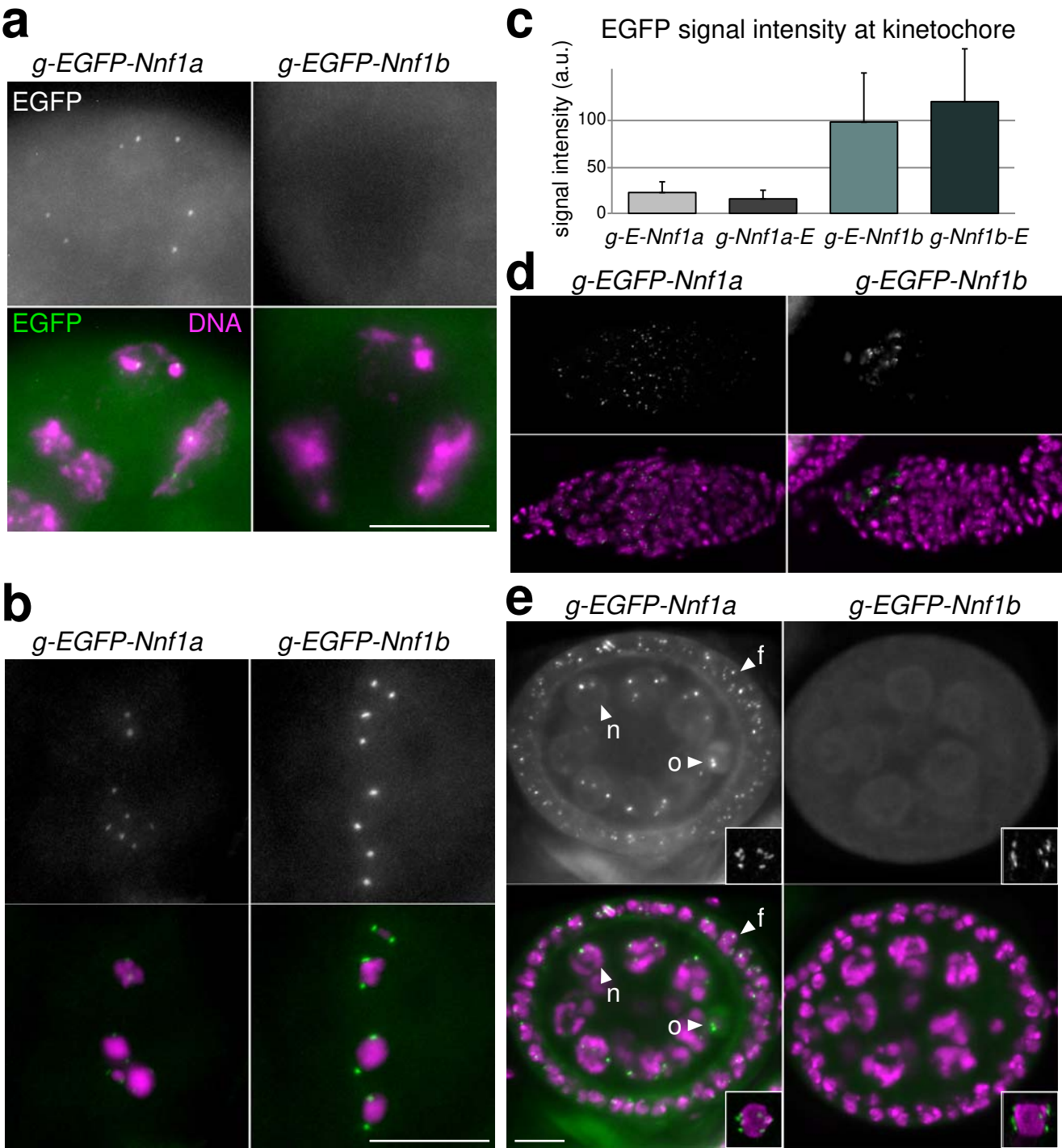


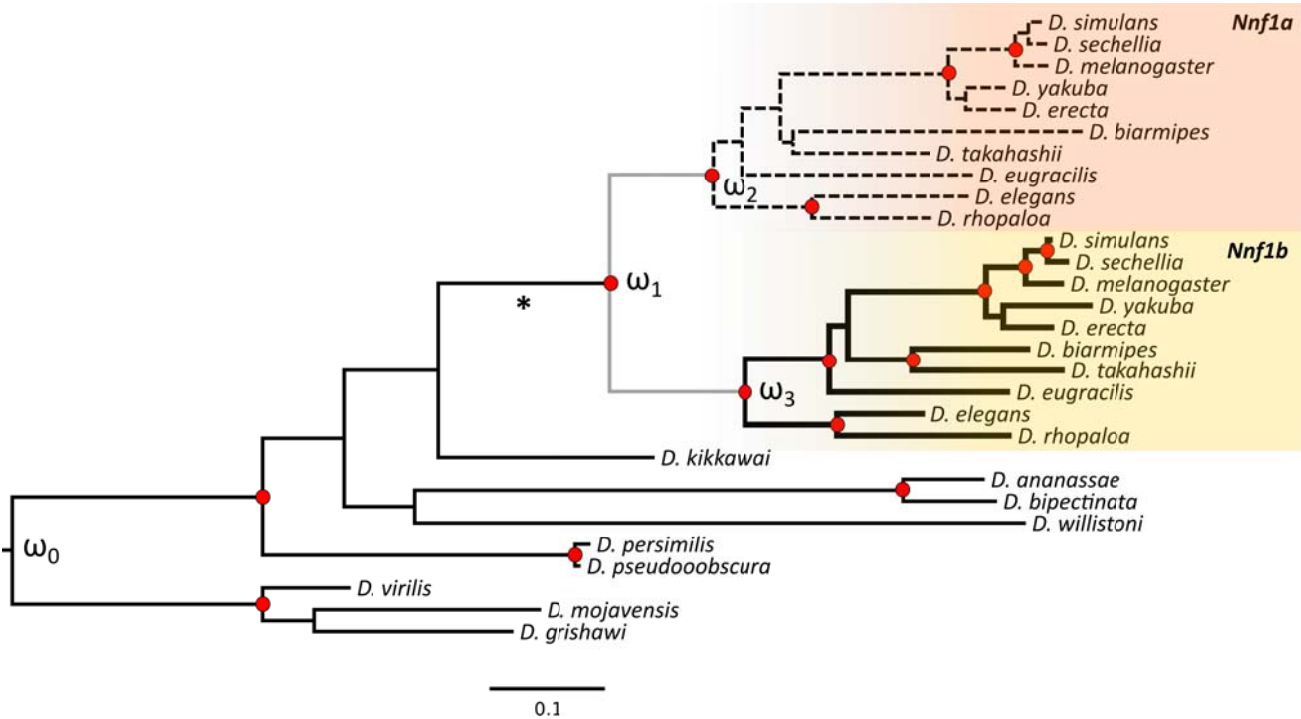
**b** EGFP signal intensity at kinetochore



**c**







Online Resource 1

***Drosophila Nnf1* paralogs are partially redundant for somatic and germline kinetochore function**

Ariane C. Blattner<sup>1)</sup>, José Aguilar-Rodríguez<sup>2,3)</sup>, Marcella Kränzlin<sup>1)</sup>, Andreas Wagner<sup>2,3,4)</sup>, Christian F. Lehner<sup>1,5)</sup>

1) Institute of Molecular Life Sciences (IMLS), University of Zurich, 8057 Zurich, Switzerland

2) Institute of Evolutionary Biology and Environmental Studies (IEU), University of Zurich, 8057 Zurich, Switzerland.

3) Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland.

4) Santa Fe Institute (SFI), Santa Fe, New Mexico 87501, United States of America.

5) author for communication: e mail: christian.lehner@imls.uzh.ch

This Online Resource 1 contains Supplementary Figure 1 and 2

DwilNnf1 1 GTAGAAGAAAGCTTTCAATCGTCATAGAATATTGTGCCACAAGTGAGGCAGGCCTATGAA  
 DmojNnf1 1 ATTGAGGAAGCTTTCCAAAGGCACAGTCCAATCGCTGCCAAAGTGAAGGGAGAATATGAA  
 DvirNnf1 1 ATTGAAGAAAGCTTTCAAAAGGCACAGCCCAATTGCGGGTAAAGTAAAGCGGAATACGAT  
 DgriNnf1 1 ATTGAGGAAGCATTTAATAGGCACAGCCCAATTGCGGCTAAAGTAAAGCGGAATATAAT  
 DbipNnf1 1 GTCGAGGCTATGTTCCGGTCACACAGGAAACAGTTGCCCAAATAAAGCAGTTCGTACGAA  
 DanaNnf1 1 GTTGATGCTATGTTCCGGCCACACAGGAAACGGTTGCTCAAATAAAGCAGGCCTACGAG  
 DpseNnf1 1 CCCGAGGCTGCTACAATCGGCACAGCCAAATTGTGCCGAGGTGAAAACGGGCCTACAG  
 DperNnf1 1 CCCGAGGCTGCTACAATCGGCACAGCCAAATTGTGCCGAGGTGAAAACGGGCCTACAG  
 DsimNnf1a 1 TCGGAAGCCGCATTTTAAACGCCACAGGGGTATTGGGCCCAAGGTGAAAACAGGCCTACGAG  
 DsecNnf1a 1 TCAGAAGCCGCATTTTAAACGCCACAGAGGTATTGGGCCAAAGGTGAAAACAGGCCTACGAG  
 DmelNnf1a 1 TCGGAAGCCGCATTTTAAACGCCACAGGGGTGTGGGACCAAAGGTGAAAACAGGCCTACGAG  
 DyakNnf1a 1 TCAGAAGCCGCATTTTAAACGCACAGGGGTGTTGGACCCCAAGGTGAAAACCTCGCCTACGAG  
 DereNnf1a 1 TCAGAAGCCGCATTTTAAACGCACAGAGGTGTTGGACCCCAAGGTGAAAACCTGGCCTACGAG  
 DtakNnf1a 1 GCAGAGGCTGCTTTTAAACGCCACAGGGCATTTGTAACCCAGGTGAAAACAGGCCTACGAA  
 DbiaNnf1a 1 ATAGAAGCTGCATTTTAAAGGCACACAGGGGATTGGTTCCGAGGTGAAGAACAGGCCTACGAG  
 DeleNnf1a 1 GCAGAAGCCGCTTTTAAACGTCACAGGGGATTGGATTACAGGTGAAGCAGGCCTACGAG  
 DrhoNnf1a 1 TCAGAAACCGCTCTTTTCAAGCGTCACAAGGGGATTGGATCCCAAGGTGAAGCAGGCCTACGAG  
 DeugNnf1a 1 ACTGAAGCTGCTTATAAACGTCACAAGGGGTATTGGACCCCAAGGTGAAGGCAAGCCTACGAT  
 DrhoNnf1b 1 ACAGAAATACGCGTTTAAACAACACGAAGCTATTATTCGCGAGGTCAATTGAGGCCTACGAT  
 DeleNnf1b 1 ACAGAGACTGCATTTTAAACGCACCAAGCTATCATTTCCCAAGGTGAAGAACAGGCCTATGAC  
 DbiaNnf1b 1 ACCGAAACCGCCTTTTAGGCGCACCAAGCGATCGTTCCCAAGGTGAAGCAGGCCTACAAG  
 DtakNnf1b 1 ACAGAAGCTGCTTTTCAAGCGTCACAGGCAATCGTGCCCAAGGTGAAGCAGGAAATACCAG  
 DsimNnf1b 1 ACCGAGGCGGCTTTCAAACGGCTCCAGGCCGTCAATTCCCAAGGTGAAGCAGGCCTACGAG  
 DsecNnf1b 1 ACCGAGGCGGCTTTCAAACGGCTCCAGGCCGTCAATTCCCAAGGTGAAGCAGGCCTACGAG  
 DmelNnf1b 1 ACCGAAGCAGCTTTCAAACGGCTCCAGGCTGTCAATTCCACAGGTGAAGCAGGCCTACGAG  
 DyakNnf1b 1 ACCGAAGCAGCTTTCAAACGGCTCCAGGCGATCATTTCCCAAGGTGAAGCAGGCCTACGAG  
 DereNnf1b 1 AGCGAAGCAGCTTTCAAACGGCTCCAGGCGATCATTTCCCAAGGTGAAGCAGGCCTACGAG  
 DeugNnf1b 1 ACCGAGTCCGCTTTTCAAGCGCACAGGCGATCGTGCCCAAGGTGAAGCAAGCTTACGAG  
 DkikNnf1 1 GTAGAAGCCGCTTTCAAGCGGCACAGAGTATCATGCCCAAGGTGAAAACAGGAGTACGAC

DwilNnf1 61 CAGGCTATAACAGATATTGCCAAGGAAGTGGAAACCTTTGCTGAAATTCCTGTCCAGCAG  
 DmojNnf1 61 AAAGCATTAATTGACTTGCCCATCTGCCTGGAGCCATTTCGAGCGATTCTATTGGAGCAC  
 DvirNnf1 61 AAAGCGCTTATGGAATTCGCATGTGCTTGGAAACCGTTTGCAGCAATTCTATTGGAGCAT  
 DgriNnf1 61 CAATCGCTCATGGAAATTCAGTGTGCTTGGAGCCGTTTGCAGCCATTCTATTGGAGCAT  
 DbipNnf1 61 GACGCCACAAAACGATGGCCGATGACTTGTTCAGCTTTTGTGTTGACATCCAGAAGGAAAAAC  
 DanaNnf1 61 GACGCCACAAAACCAATTGCCGAGATTTATCGGCGTTCGTTAACATACAAAAGGAAAAAC  
 DpseNnf1 61 GAAGCTATTTCTAAAGTTGCCGAGGATCTGGCTCCCTTTGCTGCCATTACCGCAGAGCAC  
 DperNnf1 61 GAAGCTATTTCTAAAGTTGCCGAGGATCTGGCTCCCTTTGCTGCCATTACCGCAGAGCAC  
 DsimNnf1a 61 GAAGCCATTAAACAAATCTGCGCAGATCTGCAAGCGTGGGATGCCATTTACCAGGAGCAC  
 DsecNnf1a 61 GAAGCCATTAAACAAATCTGCTCAGATATGCAAGCGTGGGATGCCATTTACCAGGAGCAC  
 DmelNnf1a 61 GAAGCCATTAAACAAATCTGCGCAGATCTGCAAGCGTGGGATGCCATTTACCAGGAGCAC  
 DyakNnf1a 61 GAAGCTATTAACAAATCTGCTCAGATTTGCAAGCGTGGGATGCCATTTACCAGGAGCAC  
 DereNnf1a 61 GAAGCGATTAAACAAATCGGCTCAGATCTGCAAGCGTGGGATGCCATTTACCAGGAGCAC  
 DtakNnf1a 61 GAAGCTATTGAAAAAATCCCTCTGATTTGGAAGCATGTGCAAGCCATTATGAGGAGCAC  
 DbiaNnf1a 61 GAGGCCATTGAAAAGACCGCTTCAACATCTGGAAGAGTGCGAAGCCATTATAGAGCAGCAC  
 DeleNnf1a 61 GAAGCCATTGGACAAATCTGTTCCGACTTGGATGCTGTGTAAGCCATTATAGAAAGTGAC  
 DrhoNnf1a 61 GAAGCCATTGGACAAATCGGTTCCGGAATTGGATGTCTTCGAAGCCATTTTTAAGGAACAC  
 DeugNnf1a 61 AAAGCCATTGGACAAATCTCCTCAGATCTCGAAGAAATCAAAGTCCCTTTACGAGAAGCAT  
 DrhoNnf1b 61 GAAGCCATTCAACAGATCTCCTCAGATCTGGATGCTGTGCTGCTATCTGGAGGAAAAAC  
 DeleNnf1b 61 GAAGTCATTCAACAGATACCTTCAGATCTGGATGTTTGTGCAAGCCATTTTGGAGGAAAAAG  
 DbiaNnf1b 61 GACGTTATTGAACAGATCCCTCAGATCTGGATTCTGCGCGCCATTCTCGAGGAGCAC  
 DtakNnf1b 61 GAAGTGATTGAACAGCTCCCTCTGATCTAGACTCTTTTGCAGCCATCATCGAGGAGCAG  
 DsimNnf1b 61 GAAGCCATTGGACAGATCTCCTCAGATCTTGAGTCGTGTGCTCCATCCTCGAGGAACAC  
 DsecNnf1b 61 GAAGCCATTGGACAGATCTCCTCAGATCTTGAGTCGTGTGCTCCATCCTCGAGGAACAC  
 DmelNnf1b 61 GAAGCCATTGGACAGATCTCCTCAGATCTTGAGTCGTGTGCTCCATCCTCGAGGAGCAC  
 DyakNnf1b 61 GAAGCCATTGGACAAATCTCCTCAGATCTTGAAATCTGCGCTCCATCCTCGAGGCGCAC  
 DereNnf1b 61 GAAGCCATTGGACAGATCCCTCAGATCTCGAGTCGTGTGCTCCATCCTCGAGGCGCAC  
 DeugNnf1b 61 GAAGCCATTGGGCTGATCCGTCAGATCTGACTCTTGTGCAACATCTTGAGGAACAC  
 DkikNnf1 61 AAGGTCATTGGACAAGTGCCTCAGACTTGCAAGCGTTCTCTGACATTCTCCTGGAGCAC

DwilNnf1	121	GACAACACTTGCTGGACACAGCAACTATTATTGAGCAGCTAAAGAATCGTATGACTGGC
DmojNnf1	121	GAGAACA CAATTCTCAACAAGGACACTTTGATTGAGCGTGTGCGCATGCGCATGAGCCAG
DvirNnf1	121	GAGAACA CTATCTTAACAAGGATACTCTAATCGAACGTGTACGCGCGAGAATGAGCCAG
DgriNnf1	121	GAGAATACAATTCTCAGCAAGGATACTTTGATCGAACGTATGCGCTCAAAAATGACCCAT
DbipNnf1	121	GGCGATGCCTACGCCGACACCGAAGCTCTGGTGCCCTAAATTGCGCGAAAAAATGACCGAC
DanaNnf1	121	GGCGATGCCTACGCCGACACCGAAGCTCTGGTGCCCAAATTGCGCGAAAAGATGACTGCC
DpseNnf1	121	GGGGACACATACTTGGGCCACCGATGAATTGGTGACTGCCATGCGCCACAAAATGACCGCC
DperNnf1	121	GGGGACACATACTTGGGCCACCGATGAATTGGTGGCTGCCATGCGCCACAAAATGACCGCC
DsimNnf1a	121	GAGCAATCGGCTCTGGACACCGAAAGCATCGTAGATCGCACTCGCAGTCTGATGACCAAG
DsecNnf1a	121	GAGCAATCGGCTCTGGACACCGAAAGCATCGTAGATCGCACTCGCAGTCTGATGACCAAG
DmelNnf1a	121	GAGCAATCGGCTCTGGACACCGAAAGCATCGTAGATCGCACTCGCAGTCTGATGACCAAG
DyakNnf1a	121	GAAAACCTCAGCTTTGGATACCGAAAGTATCGTTGATCGCACTCGTAGCCTTATGACCAAG
DereNnf1a	121	GAAAACCTCTGCTCTGGACACCGAAAGTATCGTAGATCGCACTCGCAGCCTTATGACCAAG
DtakNnf1a	121	GAGGATTCAGCTTTGGATACCGAAAGTATCATAAATCGCACTCGCTGTCTTATGACGAAG
DbiaNnf1a	121	GAAGAACTCAGCTTTGGACACAGAATATATATAAACCGCACTCGCATCTTATGACAAAG
DeleNnf1a	121	AATGAATCTGACCTTGGACACCGAAAATCTATCGAATCGCACTCGTAATCTTATGACGAAG
DrhoNnf1a	121	GAGGATTCAGCTTTAGACACCGAAAATCTTTTAAATCGCACTCGCAATCTTATGACGAAG
DeugNnf1a	121	GAGAACCTCTGATTGGACACTGAGACTATCATAAATAGCACTCGGAGTCTGATGACAAAG
DrhoNnf1b	121	GAGGATTCCTTGCTTGGACACCGACGAGATAATAAATAGCACCCGTAGGCTCATGACCAAG
DeleNnf1b	121	GAGGATTCCTTGCTTGGACACCGAACAGATGATAAATAGCACCCGTGGCTCATGACCAAG
DbiaNnf1b	121	GAAGCTCTGAGCCTGGACACGGAGAGATGGTGAACACCGCCGAAAGGTCTATGACGAAG
DtakNnf1b	121	GAATCCTCGAGCGTGGACACCGAGAAGTTGGTAAATGGCGCTCGAAAGATCATGACGAAG
DsimNnf1b	121	GAATCCACAAGCCTGGACACCGAGCAAATAGTCAGCAGCACCGGGCGGCTCATGACCAAG
DsecNnf1b	121	GAATCAACAAGCCTGGACACAGAGCAAATAGTCAGCAGCACCGGGCGGCTCATGACCAAG
DmelNnf1b	121	GAATCCACGAGCCTGGACACCGAGCAAATAATCCACAGCACCCGTGGCTTATGACCAAG
DyakNnf1b	121	GAATCCACAAGCCTGGACACAGAGCAAATAGTCAGCAGCACCGGGCGGCTCATGACCAAG
DereNnf1b	121	GAATCCACAAGCCTGGACACCGAGCAAGTAGTCAACAGCACCGGGCGGCTCATGACCAAG
DeugNnf1b	121	GAAAACCTCAAGCCTGGACACCGAGCAAATAATAAACAGCACTCGGAAGCCTTATGACGAAG
DkikNnf1	121	GAGGACTCCGTGCTGGACACAGAGGAGCTGGTAAACACCATGCGCACACAGATGACCAAC

DwilNnf1	181	ATCCTATGCGGTC TAAATCAACAA TTCTTTGACATATAATGGTGTGGA CAACAAAT TGATT
DmojNnf1	181	CTGCTGCCCAAGATCAATGAAAATTTCTTTATTTCAATGATGTGGGCAAAAAATTAATA
DvirNnf1	181	GCGTTGCCCAAGATAAACGATCATTTTTTTGTGTCCAATGATGTGGCAAGAAGTTGATT
DgriNnf1	181	GTGCTGCTCAATATATACGAGAACTTTTTTTGTTGCCACGATGTGGGCAATAAACTGATT
DbipNnf1	181	ATCATGATGAAAAATGCACCTTGCTTCTTCCACAGCAATGATATCGAAAAACAAGTTGTTG
DanaNnf1	181	ATCATGATGAAAAATGCACCTTGCTTCTTCCACAGCAATGATATCGAAAAACAAGTTGTTG
DpseNnf1	181	ATATTGAGCAAGTGAACCAGCACTTCTTCGACAGCAATGATGCAGAAAAACAAGTTGGTT
DperNnf1	181	ATATTGAGCAAGTGAACCAGCACTTCTTCGACAGCAACGATGCCGAAAAACAATTTGGTT
DsimNnf1a	181	GTGGTGCTCGAAA TGAACCGATGCTTCTTCGCCAGCAACGATGTGCCAAATAAGCTGCAA
DsecNnf1a	181	GTGGTGCTCGAAA TGAACCGATGCTTCTTCGCCAGCAACGATGTGCCAAATAAGCTGCAA
DmelNnf1a	181	GTGGTGCTCGAAA TGAACCGATGCTTCTTCGCCAGCAACGATGTGCCAAATAAGCTGCAA
DyakNnf1a	181	GTGGTGCTCGAAA TGAACCGATGCTTCTTCGCCAGCAACGATGTGCCAAATAAGCTGCAA
DereNnf1a	181	GTGGTGCTCGAAA TGAACCGATGCTTCTTCGACAGCAACGATGTGGCAAATAAGCTGCAA
DtakNnf1a	181	GCAGTTCTCAAACTGAATGGATGTTTCTTTGACAGCAATGATGTGGGAAATAAGCTAAC
DbiaNnf1a	181	GCAGTTCTCAAACTGAATGGATGTTTCTTTGACAGCAACAAATCTGGAAAAAGAGATTGCA
DeleNnf1a	181	GTGGTGCTCGAAA TGAATCGATGTTTCTTTGAAA GCAACAACTAGAAAAACAAGCTAACA
DrhoNnf1a	181	GTGGTGCTCGAAA TGAACCGATGTTTCTTTGCCAGCAACGATGTGGACAACAAGCTTACA
DeugNnf1a	181	GTGGTCTCTCGAAA TGAACCAATGTTTCTTTGAAGGCAATGATGTGGAGAACAAAGCTAGCT
DrhoNnf1b	181	ATCGTCTTGATGTAAATCAATGTTTCTTCGCCGGCAGCGATGTAAGGACTAAGTTGACC
DeleNnf1b	181	ATCGTCTTGATGTGAATCAATGCTTCTTTACCGCAACGATGTGAAGACCAAGCTGACC
DbiaNnf1b	181	ATCGTCTTAGATGTGAACGAGTGCTTCTTCGCCGGCAACGATGTGGACACCAAGCTGACC
DtakNnf1b	181	GTGCTCTCGATGTCAACCAGTGCTTCTTTGCCGGCAACGATGTGGACACCAAGCTGACC
DsimNnf1b	181	ATCGTCTCGATGTGAACCAGTGCTTCTTCTCCGGCAACGACGTGGAGACCAAGCTGACG
DsecNnf1b	181	ATCGTCTCGATGTGAACCAGTCTTCTTCTCCGGCAACGACGTGGAGACCAAGCTGACG
DmelNnf1b	181	ATCGTCTGGATGTGAACCAGTGCTTCTTCTCCGGCAACGACGTGGAGACCAAGCTGACC
DyakNnf1b	181	GTGCTCTGGATGTGAACCAGTGCTTCTTCTCCGGCAACGACGTGGAGACCAAGCTGACC
DereNnf1b	181	GTGCTCTGGATGTCAACCAGTGCTTCTTCGCCGGCAACGACGTGGAGACTAAGCTGACC
DeugNnf1b	181	ATCGTCTCGATCTTAACCAATGCTTCTTCGCTGGCAACGACGTGGAGACCAAGCTGACT
DkikNnf1	181	GTCTTGCGCAAAA TGAATCAATTCTTCTTCGACCAAGAACGATGTGGAAACAAGCTAGTG



DwilNnf1 241 ACTTTAGAGGTTCTAAAAGAAAAATTCGCGCTGCCATCAGGGCAAGAATTGGAACAACATT  
 DmojNnf1 241 ACCCTTGAGGTGCTCAAAAGAGAAAATTTGAAACCCATATAAAGGCACAAACTGGCAAAAGCTG  
 DvirNnf1 241 ACCATGGAGGTGCTTAAGGAAAAAGTTTGAAACCCATATAAAGGCACTAGCTGGAACAAACTG  
 DgriNnf1 241 ACCCTGGAGGTTCTGAAGGAAAAATTCGAGCCCCCTTAAAGCAACCAACTGGAACAATTA  
 DbipNnf1 241 GCTCTAGAAGTGCTCAAGGAAAAAGTTTCGCGGGTCAAGAGGGCAAGCAATGGAATGAAATG  
 DanaNnf1 241 GCCCTAGAAGTGCTCAAGGATAAGTTTTCGAGGTTCAGGAGGGCAAGAAATGGAATGAAATG  
 DpseNnf1 241 GTACTGGAAATGCTCAAGGAGAAGTTTCGCCCCCTACGAGGGGCACAACTGGAACGTCTG  
 DperNnf1 241 GTACTGGAAATGCTCAAGGAGAAGTTTCGCCCATACGAGGGGCACAACTGGAACGTCTG  
 DsimNnf1a 241 ACTTTGGAAATGCTCAAGGAACATTTTCGCTCCCTACGAGGGCAAGAAATGGAACACTGCA  
 DsecNnf1a 241 ACTTTGGAAATGCTCAAGGAACATTTTCGCTGGCTACGAGGGCAAGAAATGGAACACTGCA  
 DmelNnf1a 241 ACTTTGGAAATGCTCAAGGAACATTTTCGCTGCCTACGAGGGCAAGAAATGGAACACTGCA  
 DyakNnf1a 241 ACCTTGGAAATGCTCAAGGAACATTTTTCGCTGCCTACGAGGGCAAGCAATGGAACACTGCG  
 DereNnf1a 241 ACTTTGGAAATGCTCAAGGAGCATTTTTCGCTGCCTACGAGGGCAAGCAATGGAACACCGCG  
 DtakNnf1a 241 AGTTTGGAGATGCTCAAGGAACATTTTTCGCTCAATATGAGGGCAAGGATGGAATACTACC  
 DbiaNnf1a 241 ACTTTGGAGATGCTCAAGGAACAGTTTTCGCTCACTACGAAGGCAAGAAATGGAATACCGCC  
 DeleNnf1a 241 ACTTTAGAAATGCTTAAGGAGCATTTTTCGCTCCCTACGAGGGCAAGGATGGAATGCTGCA  
 DrhoNnf1a 241 ACTTTAGAGATGCTTAAGGAGCATTTTTCGCTCCCTACGAGGGCAAGGATGGAATACCGTG  
 DeugNnf1a 241 ACTCTGAGATGCTAAAGGAACATTTTTCGCTGCCTATGAGGGAAAGATGGAATGCTGCC  
 DrhoNnf1b 241 ACTTTGGAGATGCTCAAGGAGCAGTTTTCGCTCAATCACGAAGGCAAGGATGGAACACCGTG  
 DeleNnf1b 241 ACTTTGGAGATGCTGAAGGAACAGTTTTCGCTCACGAGGGCAAGGATGGAACAGGTGTG  
 DbiaNnf1b 241 ACTTTGGAGATGCTCAAGGAGCAGTTTTCGCTCCCTACAGGGCAAGAAATGGAACAGCGTG  
 DtakNnf1b 241 ACTCTGAGATGCTCAAGGAACATTTTTCGCTCCCTACGAGGGCAAGGATGGAACAGTTTG  
 DsimNnf1b 241 ACGCTGGAAATGCTCAAGGAGCAGTTTTCGCGCCACGAGGGCAAGAACTGGAACAGTCTG  
 DsecNnf1b 241 ACGCTGGAAATGCTCAAGGAGCAGTTTTCGCGCCACGAGGGCAAGAACTGGAACAGTCTG  
 DmelNnf1b 241 ACGCTGGAAATGCTTAAGGAGCAGTTTTCGCGCCACGAGGGCAAGAACTGGAACAGTCTG  
 DyakNnf1b 241 ACACCTGGAGATGCTGAAGGAACAGTTTTCGCGCCACGAGGGCAAGAACTGGAACAGCTG  
 DereNnf1b 241 ACTCTGAGATGCTCAAGGAGCAGTTTTCGCGCCCTACAGGGCAAGAACTGGAACAGCGTG  
 DeugNnf1b 241 ACTTTAGAGATGCTCAAGGAGCAGTTTTCGCTCCCTATGAGGGTACTGACTGGAAGAGTGCA  
 DkikNnf1 241 ACTTTGGAGGTTCTCAAGGAAGAGTTTTCGCGCCTACGAGGGCAAGAACTGGAATAGTCTG

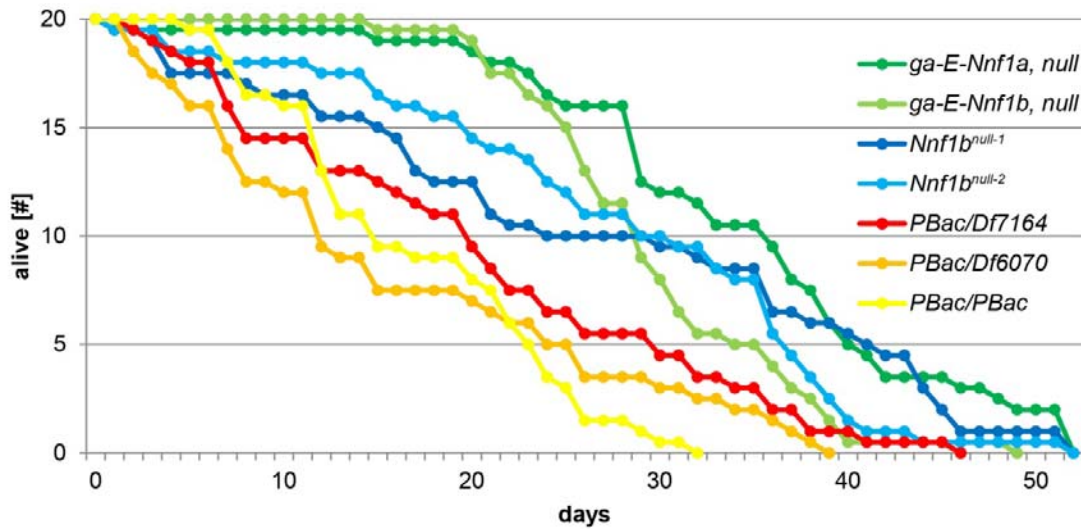
DwilNnf1 301 GCTCCCGAAGATTACACACGTCGCGGTGCGCATGCGTCTCATTTGACTCCAGCATTCGGTTT  
 DmojNnf1 301 ACACCTGAGGAAAGGACAAGACCTGTGCGCATGCGTCTTATGGATTCCAGCATTAGATT  
 DvirNnf1 301 ACGCCAGAGGAGCGGACAAAGACCGGTGCGCATGCGCTCATGGATTCCAGCATTCGATTT  
 DgriNnf1 301 ACGCCAGAGGAGCGGACAAAGACCGGTGCGCATGCGCTGATGGATTCAAGCATTCGATTT  
 DbipNnf1 301 ACCCCGAGGAGCTAACTCGTCCCTACGCATACAGCTTATGGACTCGAGCATTCGCTAT  
 DanaNnf1 301 ACTCCTGAGGAGCTAACTCGTCCATACGCATACAACTAATGGACTCCAGCATTCGCTAT  
 DpseNnf1 301 TCCCCGGAAGAAATACACGCGTCCGGTGCGCATGCGCTCTTTGGATTCCAGTATCCGCGTG  
 DperNnf1 301 TCCCCGGAAGAAATACACGCGTCCGGTGCGCATGCGCTCTTTGGATTCCAGTATCCGCGTG  
 DsimNnf1a 301 TCCCCGACGAACCTCACTCGGCCGTGCGCATGCGATTTTTGGACTTCAGCCTGGAATTC  
 DsecNnf1a 301 TCCCCGACGAACCTCACTCGGCCGTGCGCATGCGATTTTTGGACTTCAGCCTGGAATTC  
 DmelNnf1a 301 GCAACGATAAATCACTCGGCCGTGCGCATGCGATTTTTGGACTTCAGCCTGGAATTC  
 DyakNnf1a 301 TCCCCGATGAACCTCACTCGGCCGTGCGCATGCGATTTTGGACTTCAGCCTGGAATTC  
 DereNnf1a 301 TCCCCCGATGAACCTCACTCGGCCGTGCGCATGCGATTTTGGACTTCAGCCTGGAATTC  
 DtakNnf1a 301 TCCCCGGAAGCTCACTCGTCCCTTGCGCATGCGCTATATGGATTTCAGCCTTAATTTT  
 DbiaNnf1a 301 GTGCCGACAACTCACTCGGCCGTGCGCATGCGATGTTTATGATTTCAGCCTGAATTTT  
 DeleNnf1a 301 TCCCCAAAACATCTTACTCGTCCCTTGCGCATGCGGCATCTAGATTTTAGCCTGAGTTT  
 DrhoNnf1a 301 TCCCCGAAAAACTTACTCGTCCCTTGCGCATGCGACATCTAGACTTTTAGCATAGGTTT  
 DeugNnf1a 301 TCCCCGGAAGAACTTACTCGTCCGTGCGCATGCGCAGTCTAGATTTTCAAGCATACATTT  
 DrhoNnf1b 301 TCCCCGGAAGAACTCACTCGTCCGTGCGCATGCGACAGTTTGGATCTGAGCATTCGTTT  
 DeleNnf1b 301 TCCCCGGAAGAACTCACTCGTCCGTGCGCATGCGACAGTTTGGATCTGAGCATTCGTTT  
 DbiaNnf1b 301 TCCCCGAGGAACTCACCGGCCGTGCGCATGGAACAGCTTGGATTTGAGTATTCGTTT  
 DtakNnf1b 301 TCCCCGGAAGAACTCACTCGACCGGTGCGCATAAACAGTTTGGATTTGAGTATTCGTTT  
 DsimNnf1b 301 TCCCCGAGGAACTCACTCGGCCGTGCGCATGCACAACTTGAATCTAAGCATCGATTTC  
 DsecNnf1b 301 TCCCCGAGGAACTCACTCGGCCGTGCGCATGCACAACTTGAATCTAAGCATCGATTTC  
 DmelNnf1b 301 TCCCCGAGGAACTTACTCGCCCTTGCGCATGCACAACTTGAATCTAAGCATCACATTTC  
 DyakNnf1b 301 TCTCCGAGGAACTCACTCGGCCGTGCGCATGCACAACTTGGACCTGAGCATCAAGTTT  
 DereNnf1b 301 TCCCCGAGGAACTCACTCGGCCGTGCGCATGCACAACTTGGAACTAAGCATCAGATTTC  
 DeugNnf1b 301 TCCCCGGAAGACGTCACTCGGCCGTGCGCATGCACAATCTGGACCTCAGCATTCGTTTAC  
 DkikNnf1 301 AGCCCCGAGGAGCTGACACGCCGTGCGCATGCGCTCTTTGGATTCCAGCATTCGTTT

DwilNnf1 361 ATGGAACGTCAATTGGAGTCCCAACAAAGCCAATTGTTGGCCCTGACCAAGTGCCATGAG  
 DmojNnf1 361 ATACAAAAGCAGATAAAACGCACAGGAAAAAGCTATAATTGCTTGGCAAGAGAGTAAAGAA  
 DvirNnf1 361 ATCCAAAAGCAAATAGTTTCTCAGGAAAAGGCTATTATTGCTATGGCAAGAGCAGAGAG  
 DgriNnf1 361 ATTGAAAAACAGCTAAACTCTCAGGAAAAGGCTTTAATTGCTTGGCGAAGAGCAGAGAG  
 DbipNnf1 361 ATGGAGCGAAAGATAGAACTCAACAAAAGAGCTTATGGCAC TGGAAAAGAGCCCAAGCA  
 DanaNnf1 361 CTGGAGCGGAAGATAGAACTCAACAAACAGAGCTCATTGCTAC TGGAAAAGAGCAAAGCG  
 DpseNnf1 361 ATGGAGAGGGAAC TGGCAGCACAGGAAAAGGCCATTGTTGCTGTGGCCAAAAGCAAAGTT  
 DperNnf1 361 ATGGAGAGGGAAC TGGCAGCACAGGAAAAGGCCCTTGTTGCTGTGGCCAAAAGCAAAGTT  
 DsimNnf1a 361 ATGGAGCAGCAACTGGCCTCTCAGGCAAAAGAACTTATTGCTATGGCCAAAGAGCAATGCC  
 DsecNnf1a 361 ATGGAGCAGCAACTGGCCTCTCAGGCAAAAGAACTTATTGCTATGGCCAAAGAGCAATGCC  
 DmelNnf1a 361 ATGGAGCAGCAACTGGCCTCTCAGGCAAAAGAACTTATTGCTATGGCCAAAGAGCAATGCC  
 DyakNnf1a 361 ATGGAGCAGCAATTGGCCTCCAGGCAAAAGAGCTTGTTGCTATGGCCAAAAGCACTGCC  
 DereNnf1a 361 ATGGAGCAGCAGCTGGCCTCTCAGGCAAAAGAGCTTATTGCTATGGCCAAAGAGCAATGCC  
 DtakNnf1a 361 ATGGAGCAACAAC TGGAGTCTCAGGAAAAGAGCTCATTGCTATGGCCAAAAGCAAGGCT  
 DbiaNnf1a 361 GTGGAACAACAATTGAAGTCTCAGGAAAAGAACTTATTGCAACGGCCAAATACAACGAT  
 DeleNnf1a 361 ATGGATCAACAGTTGGAGTCCCAAAGAAAAGAGCTTATTGCAATGGCCAAAAGCAATGAG  
 DrhoNnf1a 361 ATGGAAGGACAATTGAAATCTCAGGAAAAGAGCTTATTGCTATGGCCAAAAGCAATGAA  
 DeugNnf1a 361 ATGGAGCGACAAC TGGAGGCTCAGGAAAAGAGCTTATCGCAATGGCCAAAAGCAATGCG  
 DrhoNnf1b 361 ATCGAGCGTCAGCTGAAAGTG CAGGAAATGGAACTTATTGCTATTTGGCAATAGTAAAGCG  
 DeleNnf1b 361 ATGGAGCGCCAAC TGAAGTCC CAGGAAAAGGAACTTATTGCTATGGCCAAAAGCAATGCG  
 DbiaNnf1b 361 ATGGAGCGCAGCTGAAGTCC CAGGAAAAGGAGCTTATTGCTATGGCGAAGAGCGTGGAA  
 DtakNnf1b 361 ATGGAGCGGCAGCTGAAGACC CAGGAAAAGGAACTTATTGCTATGGCCAAAAGCAATGAA  
 DsimNnf1b 361 ATGGAGCAACAGCTTAAGATA CAGGAAAAGGAACTCATTGCTATGGCCAAAAGCAATGAA  
 DsecNnf1b 361 ATGGAGCAACAAC TTAAGATA CAGGAAAAGGAACTCATTGCTATGGCCAAAAGCAATGAA  
 DmelNnf1b 361 ATGGAGCAACAGCTTAAGAAA CAGGAAAAGGAACTCATTGCTATGGCCAAAAGCAATGAA  
 DyakNnf1b 361 ATGGAGAAACAGCTGCAGATC CAGGAAAAGGAGTCTCATTGCTATGACCAAAAAGCAATGCA  
 DereNnf1b 361 ATGGAAGACAGCTGCAGATC CAGGAAAAGGAACTCATGGCTATGAAACAAAAGCAATCAA  
 DeugNnf1b 361 ATCGAGCGCAGCTGCAGATC CAGGAAAGAGGAGCTTATTGCTATGGCCAAAAGCAATGAA  
 DkikNnf1 361 ATGGAGAGGCAGCTGGAGGCA CAGCAAAAAGATCTCATTTCATGGCCAAAAGCAAAGCT

DwilNnf1 421 AATCGTCAAAATCTTACAAAATCTTCAAAAACGAGCGTGTCAAACCTAAAGGCCATAATGGAA  
 DmojNnf1 421 AACCGGGAACGCATTTCAAACCATTTCAAATGAACGCGTTAAAGCTTTACGCCCTGATGCAT  
 DvirNnf1 421 AATCGCGAACGTA TACAAAGCATTCAGAACGAACGCGTTAAACCTTTACGCCCTGATGCAT  
 DgriNnf1 421 AATCGGGAGCGTTTACAGAATATTCAAATGAGCGTGTGAAAACCTACGCCCTTGTGTCAT  
 DbipNnf1 421 AACCGGGAACGCTTTCAAATATATCAAACGAACGTTGTAACCTTAACGCCATAATGGAG  
 DanaNnf1 421 AACCGGGAACGCTTTCAAATATATCAAACGAACGCGTAAACCTTAATGCCATTATGGAG  
 DpseNnf1 421 AACCGAGAGCGCTTACAAAACATTCAGAACGAACGCGTTAAAGCGAACGTCTCGATGAAT  
 DperNnf1 421 AACCGAGAGCGCTTACAAAACATTCAGAACGAACGCGTTAAAGCGAACGTCTCGATGAAT  
 DsimNnf1a 421 AATCGCGAGCGACTCCAACATATCCATGACGAGCGACTGAAATGAGTGCCCAAATGGAG  
 DsecNnf1a 421 AATCGCGAGCGACTCCAACATATCCATGACGAGCGACTGAAATGAGTGCCCAAATGGAG  
 DmelNnf1a 421 AATCGCGAGCGACTCCAACATGTCCATGACAAGCGCTGAAATGACTGTCCCAAATGGAG  
 DyakNnf1a 421 CATCGTGAGCGCTTGCAAAATGTCCATGACGAGCGACTGAAATGAGTGCCCAAATGGAG  
 DereNnf1a 421 CATCGTGAGCGCTTGCAAAATATTCATGACGAGCGACTGAAATGAGTGCTATAATGGAG  
 DtakNnf1a 421 AATCGTGAGCGTATACAAAAGTACAGACGAACGACTGAAACCTTTACGCCCAAATGAG  
 DbiaNnf1a 421 AATCGTGACCGAGTACAAAATCTGCTTGACGAGCGGCTGAAACCTGAATGCCAAAATGAAG  
 DeleNnf1a 421 AATCGACAACGCC TACAAAAAATG CAGGACGAACGAGTAAACCTGAAAGCCAAAATAGAG  
 DrhoNnf1a 421 AATAGAGAACGCC TCAAGATGTG CAGAACAACGAGTAAACCTGAAAGCCAAAATAGAA  
 DeugNnf1a 421 AATCGTGAAACGCC TCAAAAATGTG CAAAGACGAGCGAGTGAAGTTACATGTCAAGATAGAG  
 DrhoNnf1b 421 AATCGACAGATTA TACGAGATGT CAAAGCGACCGTGTGAAATGGGTGTCTATGATACAA  
 DeleNnf1b 421 AATCGACAGCTTG TACGCGATGT GCAGAGCGAACGAGTGAACCTGGGCGTCATAATACAA  
 DbiaNnf1b 421 AATCGACAGCGCATTCAGGATGTGCAGGTTGAGCGCGTGAAAGTGGGACACTTAATAAAA  
 DtakNnf1b 421 AATCGACAGCGTATACATGATGTGCAAGCTGAGCGAAAGAAAGTTGAACGTTTAATAAAT  
 DsimNnf1b 421 AATCGACAGCTTATACATGATGTCCACGCTGAACGTTGAAAGTGGGATGTATGATGAAG  
 DsecNnf1b 421 AATCGACAGCTTATACATGATGTCCACGCTGAACGTTGAAAGTGGGATGTATGATGAAG  
 DmelNnf1b 421 AATCGACAGCTTATACATGATGTCCACGCTGAACGGTGAAGTGGGATGTATGATGAAG  
 DyakNnf1b 421 AATCGAAATCGTATAGATGATATCAACGCGGAACGAGTAAGAGTGGGATGTATGATGAAG  
 DereNnf1b 421 AATCGACAGCTTG TACATGATGTCCACGCGAACGAGTGAAAGTGGGATGTATGATGAAG  
 DeugNnf1b 421 AACCGTAAGCGTATACAAAGATGTACAGGCGGAGCGAGTGAAGTGGGAATTTTAAATAAAG  
 DkikNnf1 421 AATCGAGAACGCC TGCAAAACCGTG CAGAACGAACGAGTCAAACCTAACGCCCAAATGGAG

DwilNnf1	481	GAACAAATCAAGTGTATTGATGAAACA
DmojNnf1	481	CAGCAAAATGAATTACTACAGGAAATG
DvirNnf1	481	CAGCAAAACGGGGTATTACAAGGAAATG
DgriNnf1	481	CAGCAGATGGATTACTACAAGGAAATG
DbipNnf1	481	CAGCAGTTGGCGGAGTTTAATGAAATA
DanaNnf1	481	CAGCAATTGGCGGAGTTTAAGGAAATA
DpseNnf1	481	CAGAAGATAGCCAGTACGCGGAAATG
DperNnf1	481	CAGAAGATAGCCAGTACACGGGAAATG
DsimNnf1a	481	CAGCAATTGTTCGAGTACGAGAAAGTA
DsecNnf1a	481	CAGCAATTGTTCGAGTACGAGAAAGTA
DmelNnf1a	481	CAGCAATTGTTCGAGTACGAGAAAGTT
DyakNnf1a	481	CAGCAATTGTCTCAGTACGAGAAAATT
DereNnf1a	481	CAGCAATTGTCTCAGTACGATAAGGTT
DtakNnf1a	481	CAGCAATTATCTCAGTATCAGGAAATC
DbiaNnf1a	481	CAGCAATTATCTGAGTACAACTTACTA
DeleNnf1a	481	CATCAACTACTGCGAGTACCAGAAAATT
DrhoNnf1a	481	CAGCAATTATCGCAGTACCAGAAATAT
DeugNnf1a	481	GAACAAATTATCTCTATACCGGGAAATT
DrhoNnf1b	481	GAGCAGATGGCTGGGTATCAAGAAATT
DeleNnf1b	481	GAGCAGCTGGCTCAGTATCAGGAAATG
DbiaNnf1b	481	GAGAGGATGGCCAGTACCAGGAAATG
DtakNnf1b	481	GAGCGGATGGAAACAGTATCATGAAATG
DsimNnf1b	481	CAGCAGATGGCAGAGTATCAGGCCATA
DsecNnf1b	481	CAGCATATGGCAGAGTATCAGGCCATA
DmelNnf1b	481	CAGCAGTTGGCCGAGTACCAGGCCATT
DyakNnf1b	481	GAGCAGATGGCCGAGTACCAGGACATA
DereNnf1b	481	GAGCAGATGGCCGAGTACCAGGACATA
DeugNnf1b	481	GAGAGAAATGGCTCAGTATCAGGATATT
DkikNnf1	481	CAGCAGACGGCCAGTACAAGGATATT

**Supplementary Figure 1.** Multiple codon-based alignment of twenty-nine *Nnfl* nucleotide sequences. Sequences are labeled with an acronym for the species of origin (e.g. “Dsim” for *D. simulans*) followed by the name of the gene: “Nnf1” for pre-duplication orthologs and “Nnf1a” or “Nnf1b” for the two different *Nnfl* paralogs (*Nnf1a* and *Nnf1b*). For example, “DsimNnf1a” represents the sequence of the paralog *Nnf1a* from *D. simulans*. Among the twenty-nine *Nnfl* sequences are nine pre-duplication orthologs from the following species: *D. ananassae* (“Dana”), *D. pseudoobscura* (“Dpse”), *D. persimilis* (“Dper”), *D. willistoni* (“Dwil”), *D. mojavensis* (“Dmoj”), *D. virilis* (“Dvir”), *D. grimshawi* (“Dgrim”), *D. kikkawai* (“Dkik”), *D. pectinata* (“Dpec”). The remaining sequences comprise ten sequences each of the paralogs *Nnf1a* and *Nnf1b* from the following species: *D. simulans* (“Dvsim”), *D. sechellia* (“Dsec”), *D. melanogaster* (“Dmel”), *D. yakuba* (“Dyak”), *D. erecta* (“Dere”), *D. eugracilis* (“Deug”), *D. biarmipes* (“Dvir”), *D. takahashii* (“Dtak”), *D. elegans* (“Dele”), *D. rhopaloea* (“Drho”). The sequences were aligned with PRANK (Löytynoja and Goldman 2008) using as a guide tree for the alignment a recent whole genome phylogeny of *Drosophila* species (Seetharam and Stuart 2013). All sites with gaps were removed using trimAl v1.2 (Capella-Gutierrez et al. 2009). This figure was generated with BOXSHADE.



**Supplementary Figure 2.** Viability of *Nnf1* mutants. To evaluate the viability of adults of a given genotype, 10 males and 10 females were isolated 0-2 days after eclosion and pooled into a vial with standard *Drosophila* food at day = 0. Flies were transferred to a fresh vial 2-3 times per week. The vials were kept at 25°C and inspected daily to determine the number of flies alive. The average of the number of flies alive in the two replicates at the indicated time points was plotted. The analyzed genotypes were:

- *Df(2L)MK01, gDSP II.2, PBacLL02791/Df(2L)NK01, gDSP II.2, Df(2R)Exel7164; ga-EGFP-Nnf1a/+ (ga-E-Nnf1a, null)*
- *Df(2L)MK01, gDSP II.2, PBacLL02791/Df(2L)NK01, gDSP II.2, Df(2R)Exel7164; ga-EGFP-Nnf1b/+ (ga-E-Nnf1b, null)*
- *Df(2L)MK01, gDSP II.2/Df(2L)NK01 (Nnf1b<sup>null-1</sup>)*
- *Df(2L)MK01, gDSP II.2/NK01, gDSP, II.2 (Nnf1b<sup>null-2</sup>)*
- *PBacLL02791/Df(2R)Exel7164 (PBac/Df7164)*
- *PBacLL02791/Df(2R)Exel6070 (PBac/Df6070)*
- *PBacLL02791/PBacLL02791 (PBac/PBac)*